Structural and Functional Studies on βαβββ-Module Resistance Proteins from *Pseudomonas aeruginosa* PAO1

and

Structural Insights into Mycobacterial Ergothioneine Biosynthesis

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

Dissertation

von Katharina Allegra Vit
aus Bochum / Deutschland

1. Referent: Professor Dr. Wulf Blankenfeldt
2. Referent: Professor Dr. Michael Steinert
3. Referent: Professor Dr. Franz Xaver Schmid
eingereicht am: 30.03.2015
mündliche Prüfung (Disputation) am: 09.07.2015

Druckjahr 2016

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Vit, A.*, Misson, L.*, Blankenfeldt, W., Seebeck, F. P. Ergothioneine Biosynthetic Methyltransferase EgtD Reveals the Structural Basis of Aromatic Amino Acid Betaine Biosynthesis. *ChemBioChem*. **2015** Jan 2; 16(1): 119-125.

Goncharenko, K. V.*, Vit, A.*, Blankenfeldt, W., Seebeck, F. P. Structure of the Sulfoxide Synthase EgtB from the Ergothioneine Biosynthetic Pathway. *Angew. Chem. Int. Ed. Engl.* **2015** Feb 23; 54(9): 2821-2824.

**Diese Autoren haben in gleichem Maße zu der Veröffentlichung beigetragen.*

Tagungsbeiträge

Vit, A., Yu, S., Blankenfeldt, W.: Characterization of Glyoxalase I/Bleomycin Resistance Proteins from *Pseudomonas aeruginosa* PAO1. (Vortrag) 29th „Rabensteiner Kolleg“, Pottenstein, Germany, May 2011

Vit, A., Yu, S., Blankenfeldt, W.: Characterization of Glyoxalase I/Bleomycin Resistance Proteins from *Pseudomonas aeruginosa* PAO1. (Vortrag) 4th BZMB PhD Symposium of the University of Bayreuth, Selb, Germany, November 2011

Vit, A., Kalawy-Fansa, E., Yu, S., Blankenfeldt, W.: Characterization of Glyoxalase I/Bleomycin Resistance Proteins from *Pseudomonas aeruginosa* PAO1. (Vortrag) 20th Conference of the German Crystallographic Society (DGK), Munich, Germany, March 2012

Vit, A., Kalawy-Fansa, E., Yu, S., Blankenfeldt, W.: Characterization of Glyoxalase I/Bleomycin Resistance Proteins from *Pseudomonas aeruginosa* PAO1. (Poster) 5th BZMB PhD Symposium of the University of Bayreuth, Selb, Germany, November 2012, 2nd poster prize

Vit, A.: Breaking the Wall of Pseudomonas Phenazine Resistance. (Vortrag) Falling Walls Lab, Berlin, Germany, November 2013

Vit, A., Popp, M., Kalawy-Fansa, E., Yu, S., Blankenfeldt, W.: Characterization of βαβββ Resistance Proteins from *Pseudomonas aeruginosa*. (Vortrag) 6th International PhD Symposium of the Helmholtz International Graduate School, Helmholtz Centre for Infection Research, Braunschweig, Germany, December 2013

Vit, A., Popp, M., Kalawy-Fansa, E., Yu, S., Blankenfeldt, W.: Characterization of βαβββ-module Resistance Proteins from *Pseudomonas aeruginosa*. (Poster) 23rd Congress and General Assembly of the International Union of Crystallography (IUCr), Montréal, Quebéc, Canada, August 2014

Vit, A., Misson, L., Goncharenko, K. V., Mashabela, G. T., Blankenfeldt, W., Seebeck, F. P.: Structural Insights into Mycobacterial Ergothioneine Biosynthesis. (Poster) 5th Murnau Conference on Structural Biology, Murnau, Germany, September 2014

Vit, A., Popp, M., Kalawy-Fansa, E., Yu, S., Blankenfeldt, W.: Characterization of βαβββ-module Resistance Proteins in *Pseudomonas aeruginosa*. (Poster) 7th International PhD Symposium of the Helmholtz International Graduate School, Helmholtz Centre for Infection Research, Braunschweig, Germany, December 2014

Für

Aldo, Esterina, Kurt, Anna und Matteo

Index

Abbreviations

Å	Ångstrom (0.1 nm)
ADIC	2-amino-2-desoxyisochorismic acid
APS	ammonium persulfate
βME	β-mervapto-ethanol
bp	base pairs
BESSY	Berlin Electron Storage Ring Society for Synchrotron Radiation, Helmholtz Centre Berlin, Berlin, Germany
BSA	bovine serum albumin
CLec	C-type lectin
cPCR	colony polymerase chain reaction
Da	Dalton
ddH_2O	double destilled water
DNA	deoxy ribonucleic acid
dNTPs	deoxy nucleoside triphosphate
DPF	Dortmund Protein Facility at the Max Plank Institute for Molecular Physiology, Dortmund, Germany
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
EM	Electron Microscopy
EMBL	European Molecular Biology Laboratory
ESI-MS	Electrospray Ionization Mass Spectrometry
ESRF	European Synchrotron Radiation Facility, Grenoble, France
FMN	flavin mononucleotide
g	unit of the gravitational acceleration
GSH	glutathione
HMM	Hidden Markov model
HPLC	High-Performance Liquid Chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
LB	lysogeny broth
MES	2-(N-morpholino)ethanesulfonic acid
MPI	Max Planck Institute
MR	Molecular Replacement
MWCO	molecular weight cut-off
Ni(II)-IMAC	Ni(II)-immobilized metal ion affinity chromatography
NMR	Nuclear Magnet Resonance
OD_{600}	optical density at 600 nm

ORF	open reading frame
p.a.	*pro analisi*
PAGE	polyacrylamide gel electrophoresis
PCA	phenazine-1-carboxylic acid
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PDC	phenazine-1,6-dicarboxylic acid
PEG	polyethylene glycol
pH	*potentia hydrogenii*
Phyre^{2}	Protein Homology/analogY Recognition Engine V 2.0
Phz	phenazine
phz operon	phenazine biosynthesis operon
PYO	pyocyanin
QS	quorum sensing
rev	reverse primer/oligonucleotide
revNi(II)-IMAC	reverse Ni(II)-immobilized metal ion affinity chromatography
r.m.s.d.	root mean square deviation
ROS	reactive oxygen species
SAD	Single Anomalous Diffraction
SAH	S-adenosyl-homocysteine
SAM	S-adenosyl-methionine
SLS	Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland
SDS	sodium dodecyl sulfate
TAE	Tris Acetate EDTA buffer
TEV	tobacco etch virus
Tris	Tris(hydroxymethyl)amino methane
VOC	vicinal oxygen chelate
HZB	Helmholtz Centre Berlin, Berlin, Germany

Amino acids are abbreviated with one letter and three letter codes according to the recommendations of the *International Union of Pure and Applied Chemistry* (IUPAC) and of the *International Union of Biochemistry and Molecular Biology* (IUB).

PART I

Structural and Functional Studies on βαβββ-Module Resistance Proteins from *Pseudomonas aeruginosa* PAO1

1 Summary

Pseudomonas aeruginosa is a multi-drug resistant opportunistic human pathogen that produces virulence factors. Among those, the blue phenazine compound pyocyanin plays a crucial role in lung infection by *P. aeruginosa*. As pyocyanin induces the formation of reactive oxygen species, the bacterium must have developed some intracellular protection machinery against this compound. In previous studies, the proteins PA0803, PA1353 and PA4641 were identified as pyocyanin resistance proteins. These proteins bind pyocyanin and shield it inside the bacterial cell. They belong to the family of βαβββ-module containing proteins which are generally involved in detoxification of small toxic molecules.

In this study, 22 βαβββ-module proteins were identified in *P. aeruginosa* and classified according to highly conserved residues in their binding pockets. A selection of these proteins was characterized by X-ray crystallography, biochemical, biophysical and microbiological methods.

Furthermore, the crystal structures of three proteins of unknown function PA0880, PA1672 and PA4518 were determined. PA0880 is a protein probably involved in itaconate metabolism. However, the functions of PA0880, PA1672 and PA4518 remain elusive.

Four of the 22 βαβββ-module proteins, namely PA0803, PA1353 and PA4641 as well as PA3127, use two aromatic amino acids to bind small aromatic compounds via π-stacking interactions. Interestingly, PA3127 differs from the other three proteins as it contains an additional N-acetyltransferase domain. The crystal structure of PA3127 was determined during this study, and the protein was found to be a tetramer in small angle X-ray scattering measurements. The PA3127 tetramer binds bleomycin, a glycopeptide antibiotic produced by *Streptomyces* spp., and acetyl-CoA with affinities of 413 ± 4 nM and 523 ± 33 nM, respectively. In spot plate assays, PA3127 confers bleomycin resistance to *E. coli*, most probably by transforming bleomycin into acetyl-bleomycin and rendering it inactive. Bleomycin resistance is even increased in the presence of PA3128, encoded in an operon together with PA3127. Hence, *PA3127-PA3128* is a genomic bleomycin resistance operon in *P. aeruginosa*.

2 Introduction

Bacteria and humans have evolved together ever since. Hence, bacterial weapons and human immune system have been undergoing continuous adaptions due to evolutionary pressure. In order to circumvent the human immune defense, pathogenic bacterial species possess virulence factors, multi-drug efflux pumps and other detoxifying enzymes. These factors and proteins are especially successful in immunocompromised humans. Further problems in eradicating pathogenic bacteria are caused by emerging antibiotic resistances and the reduced amount of new antibiotics released by pharma industries. In developed countries, nosocomial infections cause severe infections and even death. One of the pathogens involved in these infections is *Pseudomonas aeruginosa.*

2.1 The Opportunistic Human Pathogen *Pseudomonas aeruginosa*

The rod-shaped gram-negative bacterium *P. aeruginosa* (Figure 1A) colonizes various habitats such as water, soil and other natural as well as artificial environments. *P. aeruginosa* is able to survive under aerobic as well as anaerobic conditions and is therefore very versatile.

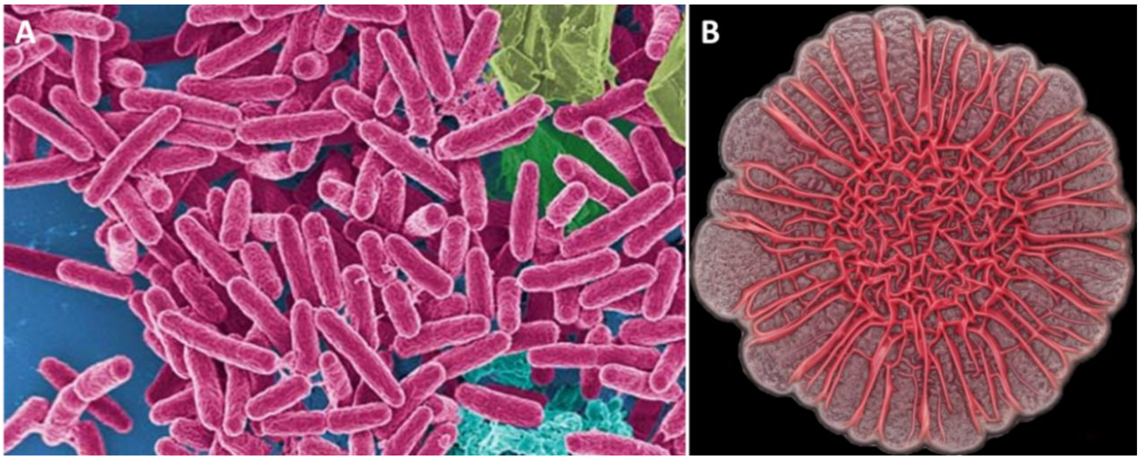


Figure 1: The opportunistic human pathogen *P. aeruginosa*. A: Colorized electron micrograph image of rod-shaped *P. aeruginosa*. B: Hyperwrinkled single colony of *P. aeruginosa*. Colonies are usually smooth and constrained, but under the absence of the respiratory pigment pyocyanin they extend to this hyperwrinkled state and form prominent spokes (Recinos *et al.*, 2012).

P. aeruginosa is an opportunistic pathogen in plants and mammals, including humans. It can cause severe infections of the respiratory tract, wounds, burns, blood and the urinary tract. Various studies have shown that these infections can have a huge impact on the infected host, leading to severe tissue damage and even to death (Nixon *et al.*, 2001; Lyczak *et al.*, 2002; Hauser *et al.*, 2011). In humans, immunocompromised patients are highly susceptible to *P. aeruginosa* infections. These infections occur either in individuals who have been treated with immunosuppressive drugs after an organ transplant or in chemotherapy, or in patients who are suffering from immune deficiencies such as AIDS. In cystic fibrosis patients, *P. aeruginosa* is able to establish chronic lung infections. Indeed, *P. aeruginosa* is the primary cause of death in cystic fibrosis-affected humans (Emerson *et al.,* 2002) and responsible for about 10% of the nosocomial infections worldwide. Hence, it is crucial to understand the physiology of this pathogen and the way of how it infects the host and establishes and maintains an infection.

The infections caused by *P. aeruginosa* are especially hard to eradicate. First, the opportunistic pathogen employs unspecific virulence factors that are used against a variety of different hosts. Upon infection of the host, the bacteria secrete virulence factors, such as exotoxin A, proteases or secondary metabolites. *P. aeruginosa* also possesses a variety of multi-drug efflux pumps that act in concert with antibiotic resistance proteins such as β-lactamases. Second, the growth of *P. aeruginosa* switches from a planktonic state to a biofilm-forming phenotype. In the biofilm, the bacteria are encapsulated in a dense matrix of extracellular polysaccharides, DNA and proteins. This matrix, together with a bacterial cell wall of low permeability, protects the bacteria from antibiotic treatment. Third, the bacteria synchronize their behavior and conceal from the immune system until they reach a certain cell density threshold. At that threshold, they simultaneously release virulence factors such as the nitrogen-containing aromatic compound pyocyanin. Pyocyanin is one of the major virulence factors of *P. aeruginosa*, also providing the characteristic blue-greenish color to the bacterium. When the virulence factors are released, *P. aeruginosa* raid the host immune defense to establish the infection.

For these reasons, the bacteria are a severe threat to human health. Continuous research in the field is needed to investigate the mechanism underlying infection and proliferation in the host and in order to develop suitable treatment for *P. aeruginosa* infections. For instance, the systems that synchronize bacterial behavior and production of virulence factors are under intensive investigation. These systems are termed the quorum sensing (QS) systems.

2.2 Synchronization of Bacterial Behavior by Quorum Sensing

A quorum sensing (QS) system is a regulatory network that is used to synchronize the behavior of a bacterial population by controlling the transcription of genes under QS control. The use of QS systems facilitates infection because the bacteria tightly suppress the production of virulence factors and hide from the host immune system. Only at a certain cell density threshold, they are able to switch on virulence factor production and flood the host with their virulence factors. This overcomes the host immune defense and often establishes a chronic infection. The most famous QS phenomenon has been observed in *Vibrio fisheri*. These bacteria accumulate in certain organs of sepia, where they live as symbionts and produce their bioluminescence after reaching a certain cell density threshold (Engebrecht *et al.*, 1983). Later, QS systems have been discovered in various bacteria, for example in the complex *Myxococcus xanthus* (Kuspa *et al.*, 1986) or in the opportunistic pathogen *P. aeruginosa* (Winson *et al.*, 1995).

Bacteria that use QS may have one ore multiple QS systems that are either arranged on the same or on multiple hierarchical levels. In the latter case, the QS systems are interconnected and their interplay is complex as they tightly regulate each other. Interestingly, *P. aeruginosa* possesses a QS network consisting of four QS systems with complex interconnections.

2.2.1 Current Understanding of Quorum Sensing in *P. aeruginosa*

The QS network of *P. aeruginosa* (Figure 2) is organized as a multi-layered hierarchy of at least four interconnected signaling systems that probably not only respond to cell density, but also to environmental stress signals (Papaioannou *et al.*, 2013; Lee & Zhang, 2014). The four QS systems provide the bacterium with flexibility to adapt to different requirements and to fine-tune its pathogenicity by regulating the production of various virulence factors. The four QS systems in *P. aeruginosa* are the *las-*, *rhl-, pseudomonas quinolone signal* (*pqs-*) and *integrated quorum sensing signal* (*iqs-*) system. Each of the four *P. aeruginosa* QS systems *las, rhl, pqs* and *iqs* requires its corresponding autoinducer molecule N-(3-oxododecanyl)-homoserine lactone (3OC$_{12}$-HSL), N-(butyryl)-homoserine lactone (C$_4$-HSL), 2-heptyl-3-hydroxy-4-quinolone, called *Pseudomonas* Quinolone

Signal (PQS), and 2-(2-hydroxyphenyl)-thiazole-4-carbaldehyde, called Integrated Quorum Sensing Signal (IQS), respectively.

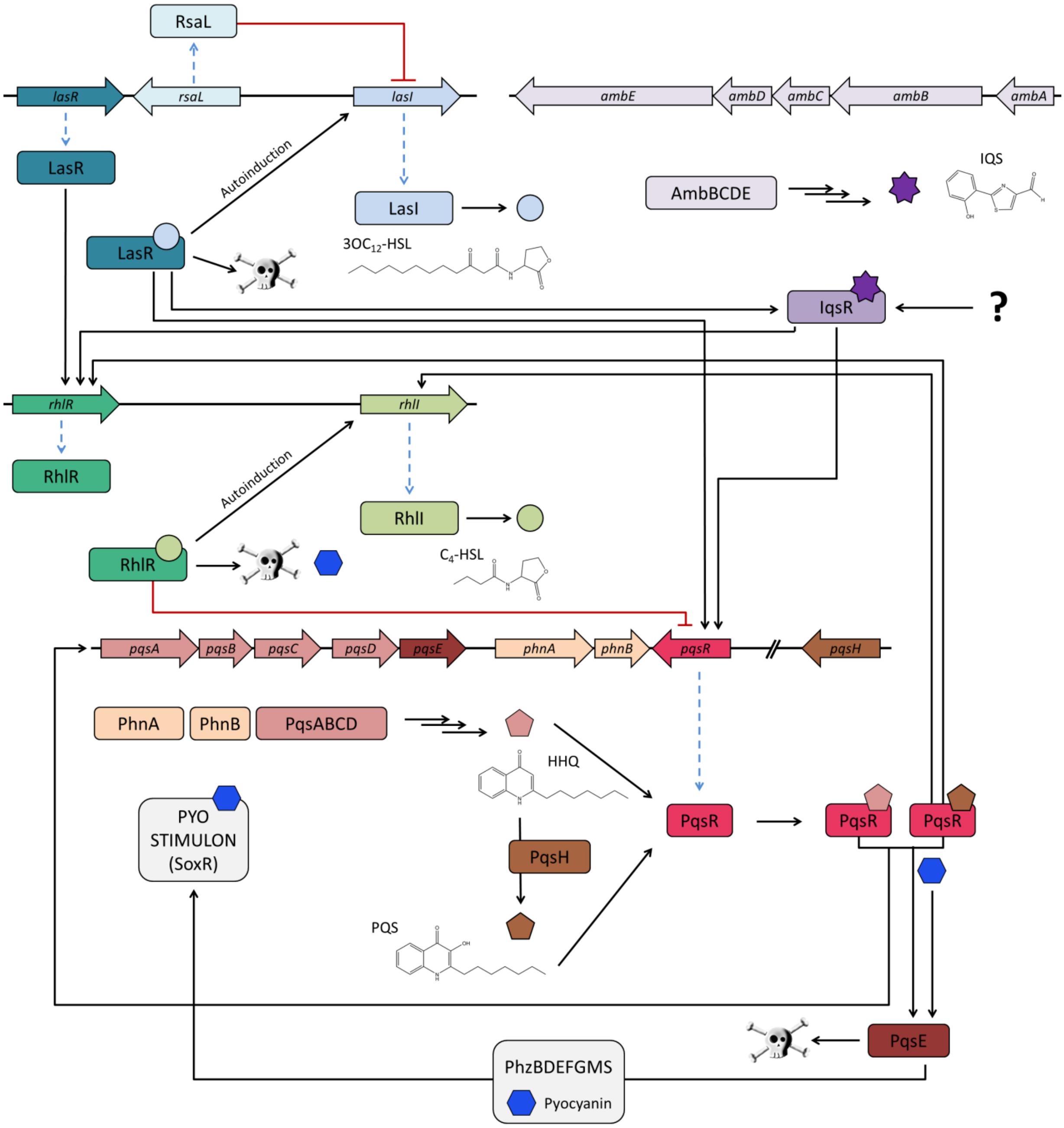


Figure 2: Simplified scheme of the four QS systems of *P. aeruginosa*. The main genes and proteins involved in the QS systems and their interconnections are presented (color code: *las*: cyan/teal; *rhl*: green; *pqs*: red/pink; *iqs*: violet). All QS systems have further connections to other enzymes or transcriptional regulators that are not depicted here. Skulls indicate the production of virulence factors such as pyocyanin. Pyocyanin (blue hexagon) and the pyocyanin-producing enzymes are under the control of the *rhl*, *pqs* and *iqs* systems. Pyocyanin itself then controls the *PYO Stimulon*, including the transcriptional regulator SoxR and the SoxR regulon. However, the exact mechanism of this regulation is not understood yet. Figure adapted from Papaioannou *et al.*, 2013; Lee & Zhang, 2014.

One example for a virulence factor that is regulated by *P. aeruginosa* QS systems is the blue phenazine compound pyocyanin. Pyocyanin production is mainly controlled by the *pqs* system, but also by the *iqs* system which probably regulates pyocyanin production through the downstream *pqs* and *rhl* QS systems. However, it is not yet clear how IQS modulates PQS and C_4HSL production (Lee *et al.*, 2013).

Interestingly, the phenazine pyocyanin is not only a virulence factor, but is a signaling molecule on its own. In fact, it is able to elicit a specific gene response and is therefore a terminal signaling factor in QS (Dietrich *et al.*, 2006). The genes activated by pyocyanin are called the *PYO Stimulon*. The *PYO Stimulon* consists of 51 genes, including the transcription factor SoxR and its regulon. 22 of these genes are upregulated, while 29 genes are downregulated in the presence of pyocyanin. Among the upregulated genes, eight genes encode putative efflux pumps and two encode transcriptional regulators (Dietrich *et al.*, 2006). Hence, exposure of *P. aeruginosa* to pyocyanin subjects dramatic effects to the gene expression profile of the bacterium, activating or deactivating several hierarchical levels of genes.

2.3 Phenazines

2.3.1 Phenazines and Phenazine Biosynthesis Operons in *P. aeruginosa*

Phenazines are small aromatic nitrogen-containing compounds. Due to their conjugated electron systems, they appear in different colors to the human eye. There are more than 160 naturally occurring phenazines known (Dictionary of Natural Products, Taylor & Francis Group, *dnp.chemnetbase.com*). Even though a huge variety of phenazine compounds is found in nature, all phenazines most probably derive from two core phenazine compounds, namely phenazine-1-carboxylic acid (PCA) and phenazine-1,6-dicarboxylic acid (PDC) (Figure 3). PCA is the precursor for asymmetric phenazines while PDC is the precursor for symmetrically modified phenazines. Phenazines are produced by a variety of bacterial species, but mainly by *Pseudomonas* and *Streptomyces*. *Pseudomonas* produces rather simple phenazines, while *Streptomyces* produce more complex and heavily modified phenazine compounds.

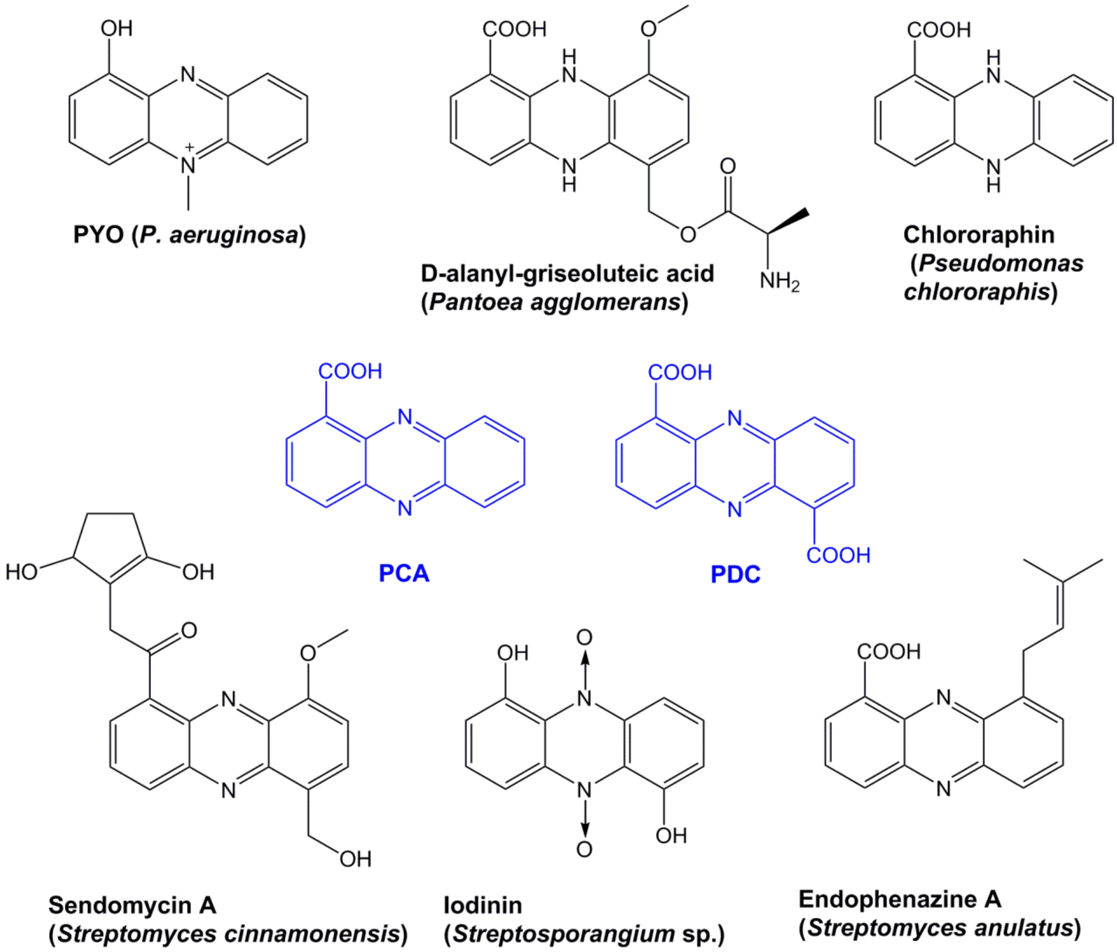


Figure 3: Naturally occurring phenazines and their origin organisms. All phenazine derivatives are probably derived from PCA and PDC which are depicted in blue. Pyocyanin (PYO) is the most abundant phenazine in *P. aeruginosa*.

In general, phenazine-producing organisms employ the proteins PhzBDEFG, encoded in the genes *phzBDEFG*, in order to produce the core phenazines PCA and PDC (Figure 4). *P. aeruginosa* produces at least four phenazines: PCA, pyocyanin, 1-hydroxy-phenazine and phenazine-1-carboxamide. However, the most abundant phenazine in *P. aeruginosa* is pyocyanin which acts as a virulence factor, signaling molecule and electron shuttle.

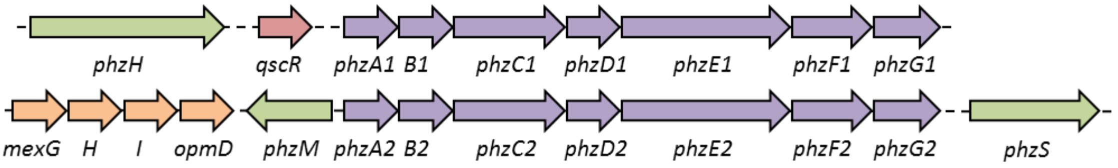


Figure 4: Phenazine biosynthesis gene clusters of *P. aeruginosa* PAO1. The core genes *phzBDEFG* are conserved in all phenazine producers. Colors indicate probable functions of the proteins (purple: core phenazine biosynthesis; green: phenazine-modifying enzymes; blue: resistance; red: regulation; orange: export proteins). Notably, *P. aeruginosa* possesses two phenazine biosynthesis (*phz*) operons which differ in architecture, expression profiles and regulation (Recinos *et al.*, 2012).

Two features of phenazine biosynthesis in *P. aeruginosa* have to be especially highlighted. On the one hand, *P. aeruginosa* possesses genes encoding transcriptional regulators such as QscR or multi-drug efflux pumps such as MexGHI-OpmD in the vicinity of its phenazine biosynthesis (*phz*) operon. On the other hand, *P. aeruginosa* possesses two *phz* operons. These two operons differ in architecture as well as in expression profiles and regulation (Recinos *et al.*, 2012). The presence of two operons is a unique feature of *P. aeruginosa*, and makes the physiology of the bacterium more complex and more difficult to study.

2.3.2 Core Phenazine Biosynthesis

The discovery of the *phz* operons in *P. aureofaciens* 30-84 (Pierson & Thomashow, 1992; Pierson *et al.*, 1995) was the first step towards detailed investigation of the phenazine biosynthesis enzymes and their mechanisms of action. Since then, most of the biosynthetic steps have been discovered and investigated and phenazine biosynthesis has been extensively studied and reviewed (Mavrodi *et al.*, 2006; Mentel *et al.*, 2009; Mavrodi *et al.*, 2010; Blankenfeldt & Parsons, 2014).

The phenazine biosynthetic pathway in *P. aeruginosa* (Figure 5) starts from chorismate which is converted to 2-amino-2-desoxyisochorismic acid (ADIC) by the ADIC synthase PhzE. The second step is catalyzed by the α/β hydrolyase PhzD and produces *trans*-2,3-dihydro-3-hydroxyanthranilic acid (DHHA) (Parsons *et al.*, 2003). DHHA is converted to highly reactive 6-amino-5-oxocyclohex-2-ene-1-carboxylic acid (AOCHC) by the enzyme PhzF (Blankenfeldt *et al.*, 2004; Parsons *et al.*, 2004b). In order to prevent AOCHC from intoxicating the bacterial cell, the enzyme PhzB assists and accelerates the condensation of two AOCHC molecules (Ahuja *et al.*, 2008). The product of the PhzB-catalyzed reaction is hexa-hydrophenazine-1,6-dicarboxylic acid (HHPDC). HHPDC can undergo isomerization or decarboxylation to tetra-hydrophenazine-1,6-carboxylic acid (THPCA). The final step is catalyzed by the flavin-mononucleotide (FMN)-dependent PhzG (Parsons *et al.*, 2004a). PhzG is able to produce either dihydroxy-PCA (DHPCA) or dihydroxy-PDC (DHPDC) depending on the number of oxidation steps it conducts on its substrate. DHPCA and DHPDC in turn undergo spontaneous oxidation to form PCA and PDC, the final products, and at the same time the precursors of all phenazine derivatives.

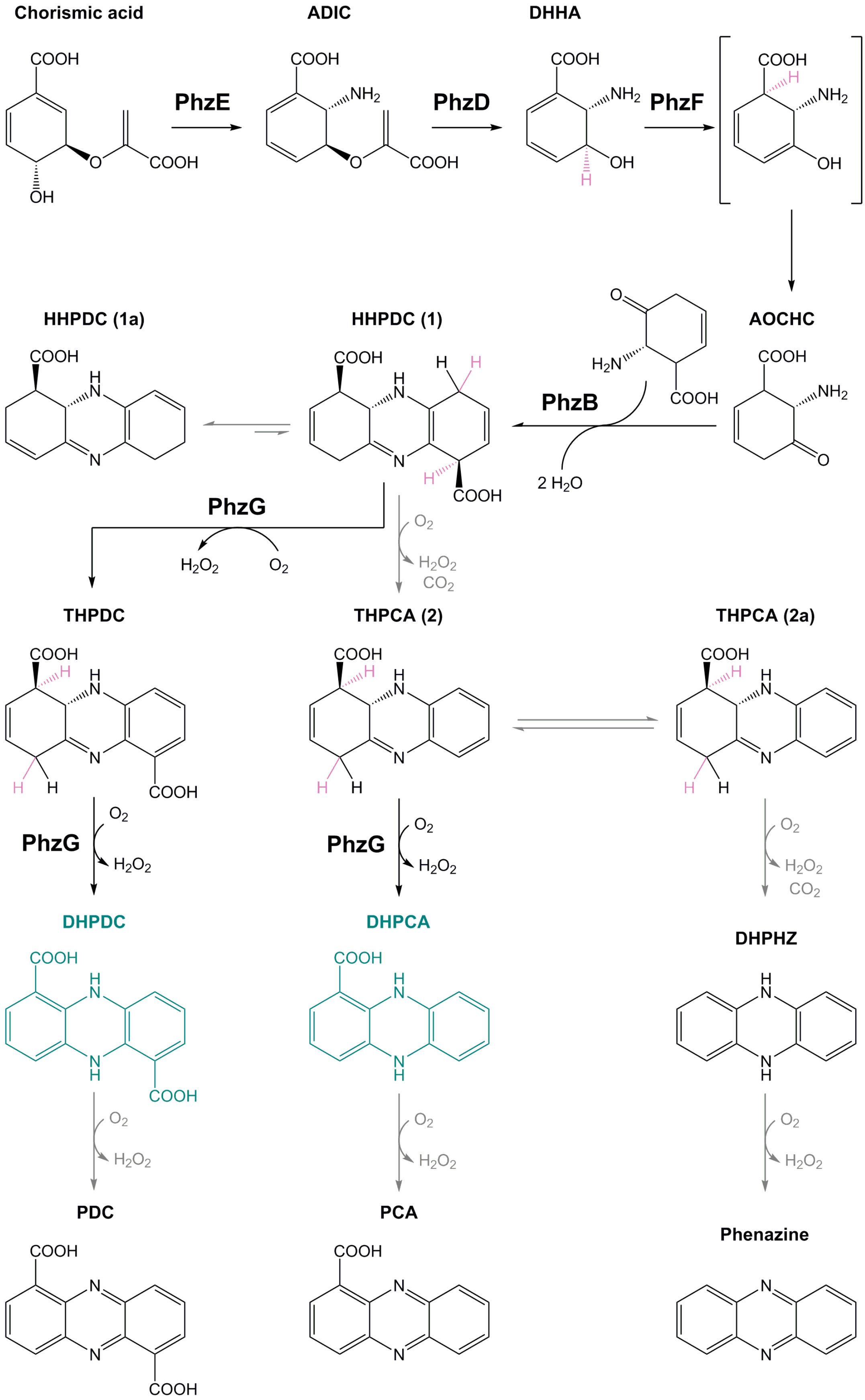


Figure 5: Current understanding of core phenazine biosynthesis. Abstracted or shifted protons are shown in pink, uncatalyzed steps are shown with grey arrows. The key products of this pathway are depicted in teal. Figure adapted from Blankenfeldt & Parsons, 2014.

2.3.3 Derivatization of PCA to Pyocyanin in *P. aeruginosa*

A unique feature of *P. aeruginosa* is the presence of two phenazine biosynthesis gene clusters. They are different in architecture (Figure 4) and it has been demonstrated that they are differently regulated (Recinos *et al.*, 2012). Removal of the *phz2* operon from the *P. aeruginosa* genome results in a more dramatic effect on the colony morphology of the bacterium than the removal of the *phz1* operon. This might be due to the presence of the *phzM* and *phzS* genes in the vicinity of the *phz2* operon. These genes encode the S-adenosyl-methionine (SAM)-dependent methyltransferase PhzM and the flavin-dependent hydroxylase PhzS. Both enzymes are essential for the conversion of PCA to pyocyanin by *P. aeruginosa* (Figure 6).

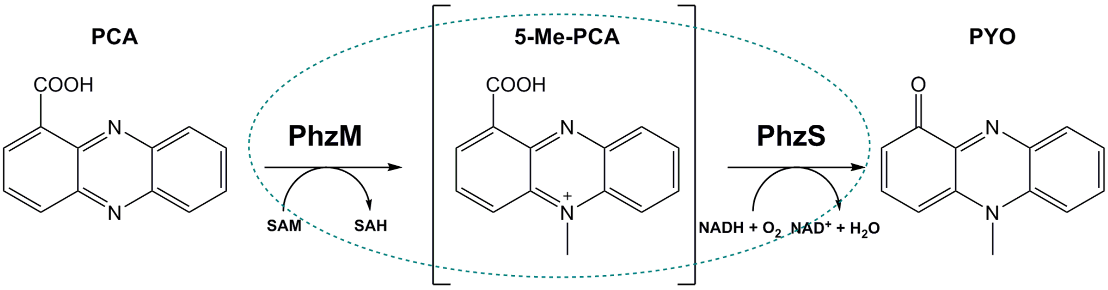


Figure 6: Pyocyanin production in *P. aeruginosa*. Both enzymes PhzM and PhzS probably form a quaternary $PhzM_2PhzS_2$ complex (indicated by the dashed ellipse). PhzM catalyzes the reaction of PCA to 5-Me-PCA, an unstable and highly reactive intermediate. This intermediate is directly transferred to PhzS. PhzS then catalyzes the reaction of 5-Me-PCA to pyocyanin.

2.3.4 Pyocyanin – Properties, Functions and Toxic Effects

The color of the most abundant phenazine in *P. aeruginosa*, pyocyanin (5-Methylphenazin-1-one), ranges from pink at acidic pH, to green at a pH of 5 and blue at basic pH. Pyocyanin (Figure 7) can even be colorless in its reduced form. The compound imparts the specific blue-greenish color to the bacterium in stationary growth phase or in biofilms. This blue-greenish color is also visible in the sputum of individuals affected by a *P. aeruginosa* lung infection, where it is found in up to 100 µM concentration (Wilson *et al.*, 1988; O'Malley *et al.*, 2004). In addition to performing multiple tasks in *P. aeruginosa*, pyocyanin is a virulence factor and plays a crucial role during host infection. Pyocyanin promotes virulence by interfering with several cellular functions in host cells including electron transport, cellular respiration, energy metabolism, gene expression, and innate immune mechanisms (Rada & Leto, 2013). For instance, pyocyanin was demonstrated to play a crucial role in lung pathogenesis (Caldwell *et al.*, 2009), but *P. aeruginosa* mutants lacking pyocyanin were attenuated in both acute and chronic lung infection models in mice (Lau *et al.*, 2004).

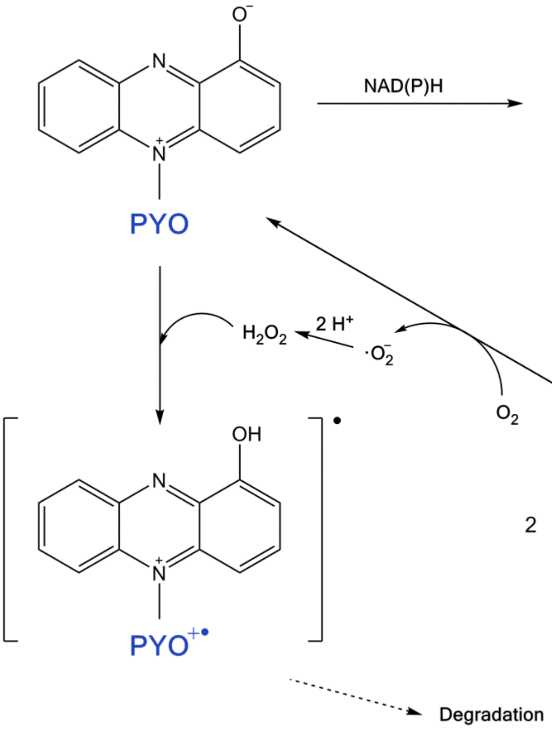


Figure 7: Redox-activity and electron shuttle properties of pyocanin from *P. aeruginosa* PAO1. PYO reacts with NAD(P)H to yield PYOH$_2$. PYOH$_2$ reduces O_2 to O_2^- or reacts with PYO to form two PYOH$^\bullet$ radicals. PYOH$^\bullet$ reduces O_2 to O_2^- . Figure adapted from Reszka *et al.*, 2004.

The basis for the multifaceted role of pyocyanin is the compound's fundamental ability to alter redox homeostasis, increase oxidative stress and induce apoptosis (Jacob *et al.*, 2011). It appears that pyocyanin is able to directly generate reactive oxygen species (ROS) without assistance of the mitochondrial respiratory chain (O'Malley *et al.*, 2004; Reszka *et al.*, 2004). Pyocyanin reacts with reducing agents such as NAD(P)H to yield hydrophenazine (PYOH$_2$). Then PYOH$_2$ primarily reduces O_2 to H_2O_2, but may also react with a second pyocyanin molecule to form two radicals (PYOH$^\bullet$), which can reduce O_2 to $O_2^{-\bullet}$ (Reszka *et al.*, 2004). Additionally, pyocyanin seems to accept one electron from glutathione (GSH), thus using the cell's own antioxidant defense to cause oxidative stress (O'Malley *et al.*, 2004). Hence, pyocyanin can adapt different oxidation states and possesses zwitterionic properties. As a zwitterion, pyocyanin might be able to easily penetrate the cellular membrane and enter a host cell. Inside the host cell, the oxidative effect of pyocyanin increases the innate immune response, for instance by upregulation of interleukin-8 expression (Denning *et al.*, 1998b; Look *et al.*, 2005). In neutrophils, ROS generation by pyocyanin and decrease in cyclic adenosine monophosphate (cAMP) concentration induce apoptosis and kill these cells (Usher *et al.*, 2002). Furthermore, pyocyanin is able to alter calcium signaling. Since Ca^{2+} is involved in regulation of ion transport, mucus secretion and ciliary beat, disruption of Ca^{2+} homeostasis might be an additional explanation for the importance of pyocyanin in lung pathogenesis caused by *P. aeruginosa* (Denning *et al.*, 1998a). Interestingly, pyocyanin is important for chronic lung infection in mice and long-term administration of pyocyanin alone is able to induce a cystic fibrosis-like lung phenotype (Caldwell *et al.*, 2009; Hao *et al.*, 2012). Taken together, pyocyanin alters several cellular functions and promotes virulence of *P. aeruginosa* in a broad range of target organisms, but especially in immunocompromised humans. The abilities of *P. aeruginosa* to interfere with redox homeostasis and signaling, gene regulation and energy metabolism of respiratory epithelial cells have recently been reviewed in detail (Rada & Leto, 2013).

In *P. aeruginosa*, pyocyanin acts as a mobile electron carrier and helps *P. aeruginosa* to survive under oxygen-poor conditions by accepting and transporting electrons produced in respiration away from the bacterium to acceptors in remote places (Rada & Leto, 2013). First evidence was found when *P. aeruginosa* mutants lacking pyocyanin formed large wrinkled colonies compared to the wild type’s smooth and small colonies (Dietrich *et al.*, 2008). This demonstrates the function of pyocyanin as a respiratory pigment and confirms its ability to modulate the structural organization of cellular communities. Later, the *pyocyanin electrocline* was discovered, a redox-potential gradient that can reach up to 400 µm beyond the biofilm surface (Koley *et al.*, 2011). This electrocline is formed under electron-acceptor limiting conditions, and is also observable in planktonic growth state as a color gradient form oxidized blue pyocyanin to reduced colorless pyocyanin (Rada & Leto, 2013). The phenazine-facilitated electron transfer was shown to promote anaerobic survival of *P. aeruginosa* (Wang *et al.*, 2010). Interestingly, growth conditions favoring the formation of the electrocline correlate to an increase in soluble iron (Koley *et al.*, 2011). In fact, pyocyanin is able to increase the bioavailability of ferric (hydr)oxides (Wang & Newman, 2008).

In contrast to its negative impact on host cell metabolism, gene regulation and redox homeostasis, pyocyanin does interfere in a beneficial way with *P. aeruginosa* cells, promoting survival of the bacterium. Most notably, the compound is able to regulate the gene expression profile of *P. aeruginosa*, since pyocyanin is a signaling molecule in QS and regulates the *PYO Stimulon* (Section 2.2.1), including eight transmembrane efflux pumps and the transcriptional regulator SoxR.

2.4 Phenazine Resistance

Phenazines and especially pyocyanin are reversibly redox-active metabolites able to generate reactive oxygen species. Hence, they act as virulence factors and are toxic to any living organism. For this reason, phenazine-producing bacteria must have developed resistance mechanisms against their self-produced phenazines. These defense responses or resistance mechanisms can be of rather general type, or they can be pyocyanin-specific resistance factors. Recent findings indicate that *P. aeruginosa* possesses general as well as specific defense mechanisms against pyocyanin and pyocyanin's action.

2.4.1 General Mechanisms of Phenazine Resistance in *P. aeruginosa*

Little is known about the defense response or resistance mechanisms to pyocyanin of *P. aeruginosa itself.* However, it is known that the bacterium is insensitive to the toxic effects of pyocyanin and that pyocyanin does not affect the growth rate of the bacterium (Baron & Rowe, 1981). In contrast, the growth rate of *E. coli* is dramatically reduced upon pyocyanin exposure (Hassett *et al.*, 1992). This effect is potentially caused by a more efficient uptake of exogenous pyocyanin by *E. coli*, and an increased formation rate of pyocyanin radicals and ROS in *E. coli*. In contrast, *P. aeruginosa* forms markedly less radicals when exposed to pyocyanin (Hassett *et al.*, 1992). Interestingly, NAD(P)H is capable to reduce pyocyanin in a non-enzymatic reaction. *P. aeruginosa* maintains a four-fold lower NAD(P)H-level than *E. coli*, and it is tempting to speculate that this explains the decreased level of pyocyanin oxidation in this bacterium. In fact, upon addition of *P. aeruginosa* cell extracts, the velocity of the non-enzymatic reduction of pyocyanin by NAD(P)H is reduced by two-fold. The reason for this remains elusive. In contrast, *E. coli* cell extract accelerates the reaction due to the presence of a NAD(P)H oxidoreductase (Hassett *et al.*, 1992). Under low-phosphate conditions, pyocyanin production levels are elevated, and *P. aeruginosa* produces high levels of multiple catalases (Hassan & Fridovich, 1980; Hassett *et al.*, 1992). A two-fold increase in catalase activity is also observed upon exposure to exogenous pyocyanin, while the absence of pyocyanin in a pyocyanin-deficient *P. aeruginosa* mutant is associated with 30% less catalase activity than in the wild type pathogen (Hassett *et al.*, 1992).

Additionally, *P. aeruginosa* produces two metal-dependent superoxide dismutases (SOD), an Fe(II)- and a Mn(II)-SOD, to detoxify ROS. The activity of both SODs and the catalases is increased 5-fold under low phosphate conditions (Hassett *et al.*, 1992). SOD activity in *P. aeruginosa* is also increased upon exposure to pyocyanin (Hassett *et al.*, 1992). Later, the expression of catalase- and SOD-encoding genes was found to be under the control of QS (Hassett *et al.*, 1999). In hindsight, the effect of pyocyanin on the expression profile of *P. aeruginosa* is not very surprising, as pyocyanin is a signaling molecule in QS (Dietrich *et al.*, 2006) that regulates its own set of target genes, the *PYO Stimulon*.

Taken together, the expression of catalases and SODs in *P. aeruginosa* is a rather a general response to oxidative stress, as it occurs also upon exposure to other ROS-generating molecules (Hassett *et al.*, 1992). Still, there is less intracellular pyocyanin-mediated redox cycling in *P. aeruginosa* than in *E. coli*. Furthermore, *P. aeruginosa* cell extracts probably contain soluble factors not present in *E. coli* which may contribute to its resistance to pyocyanin-mediated oxidative stress and pyocyanin itself (Hassett *et al.*, 1992). Hints to such factors are found in the *phz* operons of phenazine producers other than *P. aeruginosa*, for instance the non-pathogenic phenazine producer *Pantoea agglomerans*.

2.4.2 Specific Phenazine Resistance in *P. agglomerans*

In general, compound-specific resistance can occur in two different ways: either the toxic compound is destroyed by an enzyme, for instance as ampicillin is destroyed by β-lactamase; or it is bound and shielded by a chaperone until it is secreted by export machinery. The former is not very likely in the case of phenazines, since the bacterium invests energy to produce phenazines, and subsequent destruction of these would in turn be a waste of energy. Hence, phenazine-producers might use a phenazine resistance protein which acts as a binding protein or chaperone towards phenazines. However, no evidence for such a chaperone-like protein can be found in the *phz* operons of *P. aeruginosa.*

Interestingly, such a phenazine resistance protein is found in the *phz* operon of the non-pathogenic bacterium *P. agglomerans* (previously known as *Enterobacter agglomerans* or *Erwinia herbicola*). The *P. agglomerans phz* operon comprises 16 open reading frames (ORFs) (Figure 8A, Giddens *et al.*, 2002). Among these genes, five encode the essential core phenazine biosynthesis enzymes PhzBDEFG, and nine additional genes encode the phenazine modifying enzymes EhpFGHIKLMNO which catalyze the modification of PCA to D-alanyl-griseoluteic acid (AGA, Figure 3), the phenazine derivative produced by *P. agglomerans.* Additionally, the *phz* operon of *P. agglomerans* includes two non-biosynthetic genes (Figure 8A). The *ehpJ* gene was identified to encode an α-helical transmembrane protein that is potentially involved in AGA transport across the membrane of *P. agglomerans*. However, no fold or function could be assigned to the 16th gene of this operon. Notably, disruption of this gene by transposon insertion resulted in loss of AGA resistance in *P. agglomerans* (Giddens *et al.*, 2002). In agreement with this result, the removal of *ehpR* from *P. agglomerans* during the creation of a *P. agglomerans* ΔAGA yields a strain with similar sensitivity to AGA as *E. coli* (Giddens *et al.*, 2002). On the basis of these observations, this 16th gene was termed *ehpR* (*Erwinia herbicola* phenazine resistance protein). The potential phenazine resistance factor EhpR encoded by this gene was characterized in order to analyze the resistance mechanism mediated by EhpR on the molecular level. For this purpose, the crystal structure of this protein was determined in complex with griseoluteic acid (GA) (Figure 8B and C, Yu *et al.*, 2011).

The AGA resistance protein EhpR from *P. agglomerans* belongs to the βαβββ-module containing protein family or glyoxalase I/bleomycin resistance proteins. These proteins form obligate dimers and possess two ligand binding sites at the dimer interfaces. In fact, the GA binding site is located at the dimer interface (Figure 8C). EhpR is able to bind GA between two aromatic residues, Y43 and F109*, each belonging to one protein of the EhpR dimer. Binding of GA is mainly mediated through π-stacking interactions, but GA is also forming hydrogen bonds with R42 and W56. EhpR probably shields (A)GA from the intracellular space and prevents it from unraveling its toxic effects. EhpR does not alter

(A)GA chemically, but probably transports it to its export machinery EhpJ. The dissociation constant K_D of GA and EhpR was determined as 244 ± 45 µM using microscale thermophoresis (Yu *et al.*, 2011). Even if this is a rather high K_D, it confirms that EhpR acts as a molecular chaperone for (A)GA. The relatively low affinity of (A)GA to EhpR may be required for their molecular function and for the release of AGA to the extracellular space.

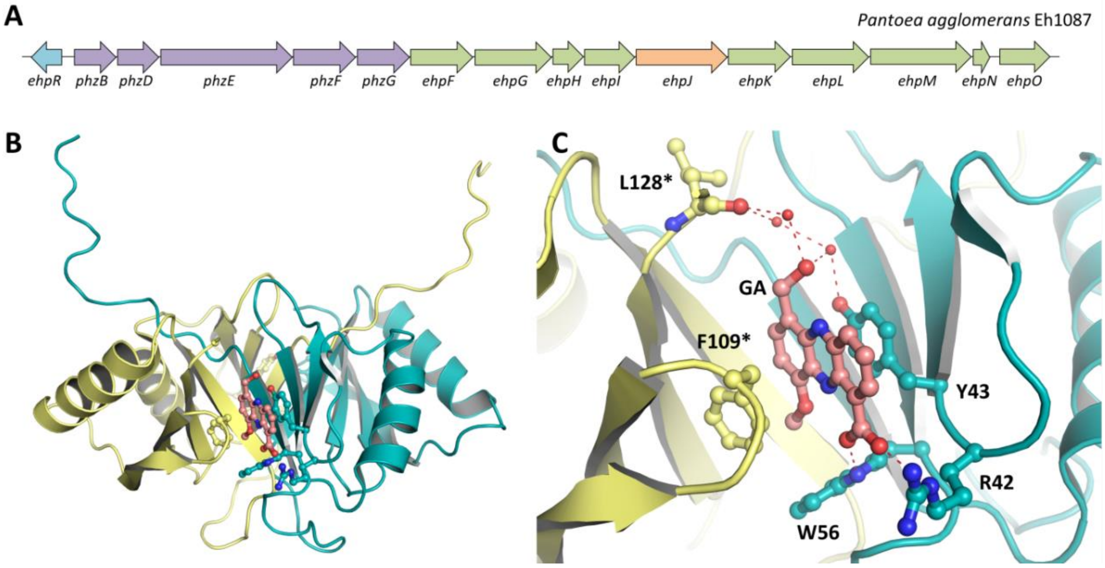


Figure 8: The βαβββ-module protein EhpR confers resistance against self-produced phenazines to *P. agglomerans*. A: D-alanyl-griseoluteic acid biosynthesis operon of *P. agglomerans*. Core phenazine biosynthesis genes are shown in violet, phenazine-modifying enzymes in green, the probable transporter in orange and the resistance gene *ehpR* in blue. B: EhpR in complex with griseoluteic acid (GA) at 1.0 Å resolution (Yu *et al.*, 2011). The protein is a dimer with two GA binding sites at the dimer interfaces. C: Phenazine binding site of EhpR with GA. GA is sandwiched between two aromatic residues Y43 and F109 via π-stacking interactions. R42 and W56 are engaged in hydrogen bonds with GA.

Together, phenazine binding proteins like EhpR and efflux pumps like EhpJ might act in concert to protect the cell from the phenazines' action and export them to extracellular space. The function of EhpR may be applicable to other aromatic or toxic compounds which could be shielded and cleared from the cell in this manner, especially in bacteria. Therefore it is interesting to examine the molecular architecture this protein and search the genome of *P. aeruginosa* for EhpR-like proteins.

2.4.3 Evidence for Specific Pyocyanin Resistance in *P. aeruginosa* PAO1

No evidence for an EhpR-like protein is found in the vicinity of the *phz* operons of *P. aeruginosa*. Due to this fact, one may assume that either *ehpR*-like genes are not present in

P. aeruginosa, or they might be located elsewhere in the genome. For bacteria, the latter does not seem very obvious or convenient, since it is thought that proteins fulfilling a function together are encoded together in the same operon. Still, they might be located elsewhere, and regulated by QS systems or other transcriptional factors.

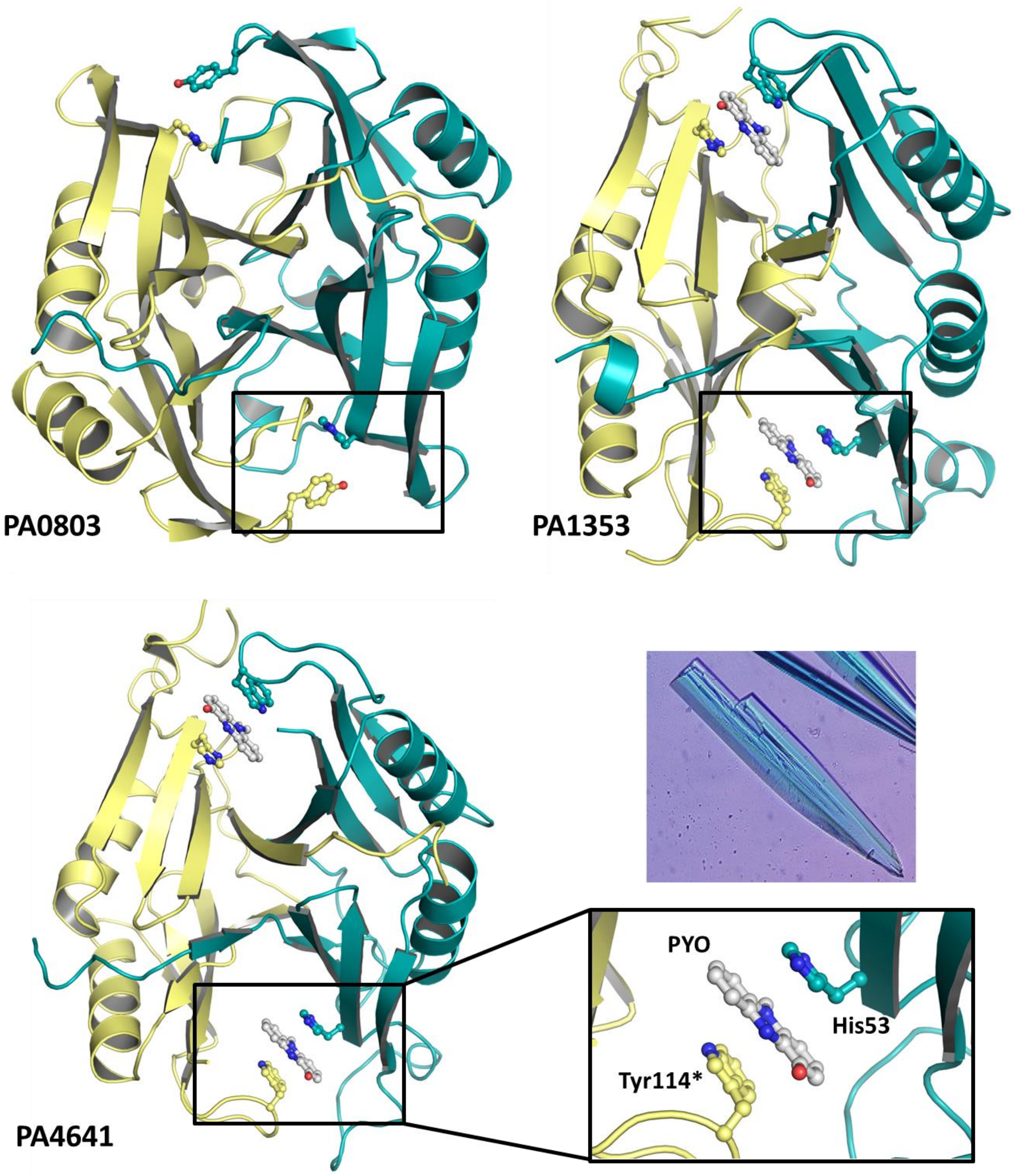


Figure 9: Crystal structures of pyocyanin-binding βαβββ-module resistance proteins PA0803, PA1353 and PA4641 from *P. aeruginosa* PAO1 (Yu, 2009; Kalawy-Fansa, 2010). The proteins form obligate dimers (monomers shown in yellow and teal). They harbor one pyocyanin binding site at each dimer interface. In PA0803, Tyr122* would have to undergo a conformational change for the protein to bind pyocyanin between residues His49 and Tyr122*. In PA1353 and PA4641, pyocyanin (grey) is bound via π-stacking interactions between residues His53 and Tyr114* (Kalawy-Fansa, 2010). Crystals of PA4641 soaked with 1 mM pyocyanin are pigmented by the compound (Kalawy-Fansa, 2010).

Notably, a *BLAST* search with the EhpR sequence identified a small protein encoded by the ORF *PA0803* in *P. aeruginosa* PAO1. The sequence identity between these proteins is relatively low (14%), but PA0803 was predicted to have the same fold as EhpR. Indeed, this prediction was confirmed when its structure was determined (Figure 9, Yu, 2009). As soon as PA0803 was identified, a *Dali* search using the structure of PA0803 on the *P. aeruginosa* proteome identified PA1353 and second *BLAST* search with the PA1353 sequence in turn identified PA4641. Again, both proteins share low sequence similarity with PA0803. In 2010, the crystal structures of PA1353 and PA4641 were both determined as apo forms and in complex with pyocyanin (Figure 9, Kalawy-Fansa, 2010).

The three proteins PA0803, PA1353 and PA4641 from *P. aeruginosa* PAO1 share structural similarity with EhpR from *P. agglomerans.* The proteins form homodimers and possess two pyocyanin binding sites which are located at the dimer interfaces. In these centers, the proteins employ two aromatic residues, one residue from each protein chain of the homodimer, for pyocyanin binding which is accomplished by π-stacking interactions. PA0803 employs residues His49 and Tyr122* (* indicating a residue from the second protein of the dimer), while PA1353 and PA4641 use His53 and Tyr114* for pyocyanin binding. In isothermal titration calorimetry (ITC), pyocyanin binding constants have been determined as 3.92 ± 0.22 µM for PA0803 (Yu, 2009), as 3.40 ± 0.25 µM for PA1353 and as 0.64 ± 0.35 µM for PA4641 (Kalawy-Fansa, 2010). These findings suggest that *P. aeruginosa* possesses pyocyanin resistance proteins similar to EhpR from *P. agglomerans*.

2.5 βαβββ-Module Containing Resistance Proteins

The GA resistance protein EhpR from *P. agglomerans* as well as the potential pyocyanin resistance proteins PA0803, PA1353 and PA4641 from *P. aeruginosa* PAO1 belong to the clan of glyoxalases (*Pfam* CL0104), also called vicinal oxygen chelate (VOC) superfamily (Armstrong, 2000) or βαβββ-module containing protein superfamily (Figure 10). This superfamily includes glyoxalase I, II, III, IV and V proteins as well as YecM proteins, CCpA-N and 3-demethylubiquinone-9 3-methyltransferases. However, all of these proteins are α/β class proteins (Levitt & Chothia, 1976), very similar in three-dimensional architecture and are predominantly found in bacteria, but also in eukaryotes and archaea (Finn *et al.*, 2014). Members of the βαβββ-module containing protein family share low sequence identity (usually <20%) but are very versatile regarding their function. The basic βαβββ-module spans 50-60 amino acid residues. Each module is arranged in a sequential order of $\beta_1\alpha\beta_2\beta_3\beta_4$, but in a spatial order and a strand polarity of $\beta_1^{+}\alpha\beta_4^{+}\beta_3^{-}\beta_2^{+}$ (Bergdoll *et al.*, 1998).

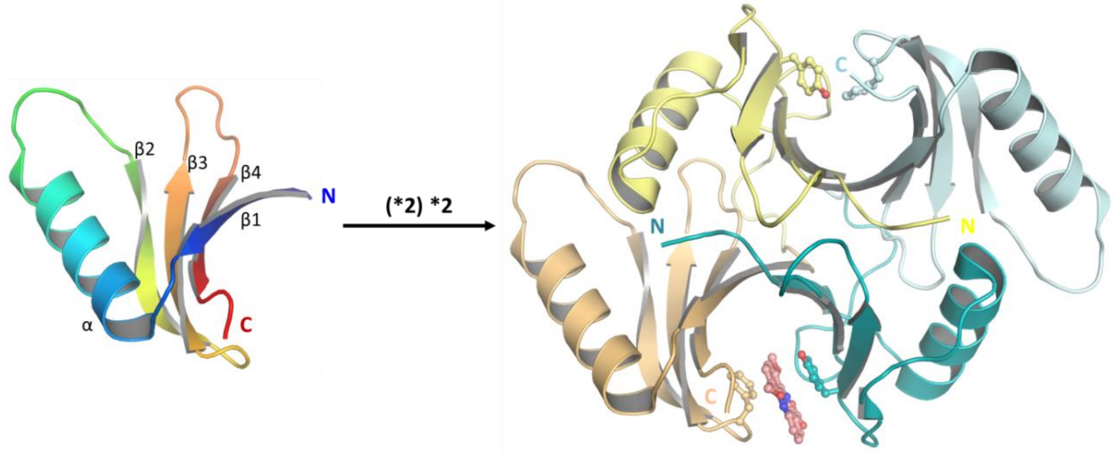


Figure 10: General architecture of glyoxalase I/bleomycin resistance proteins or βαβββ-module containing proteins, using EhpR (Yu *et al.*, 2011) as an example. On the left side, the βαβββ-module is depicted. Four of these modules are needed to form a fully functional protein. A functional protein is either encoded as one protein composed of four modules, or as a protein with two modules that forms stable homodimers.

A fully functional βαβββ-module protein is composed of four superimposable modules. These four modules can be either encoded in one protein chain or two modules can be in encoded in two protein chains each. Hence, βαβββ-module protein containing two modules are obligate dimers. However, dimerization of βαβββ-module containing proteins can be achieved in two different ways (Bergdoll *et al.*, 1998). The possible assemblies are illustrated in Figure 11 using 2,3-dihydroxy biphenyl 1,2-dioxygenase (DHBD) from *Burkholderia xenovorans* BL400 (PDB entry 1HAN, Han *et al.*, 1995), human glyoxalase I (Glo1) (PDB entry 1FRO, (Cameron *et al.*, 1997) and RHA04853 (PDB entry 2R6U, Zhang *et al.*, to be published).

In all quaternary assemblies of βαβββ-module proteins, the β_1 strands of each module are interacting to form an extended β-sheet structure (Bergdoll *et al.*, 1998). In the first possible dimerization modus, two neighboring modules 1 and 2 can dimerize with other two modules **3** and **4** via back-to-back contacts. By doing so, the α-helices of modules **1** and **3** and modules **2** and **4** are interacting with each other. Hence, spatial interaction of α-helices occurs between modules which are not neighboring in the protein sequence (Figure 11, DHBD monomer).

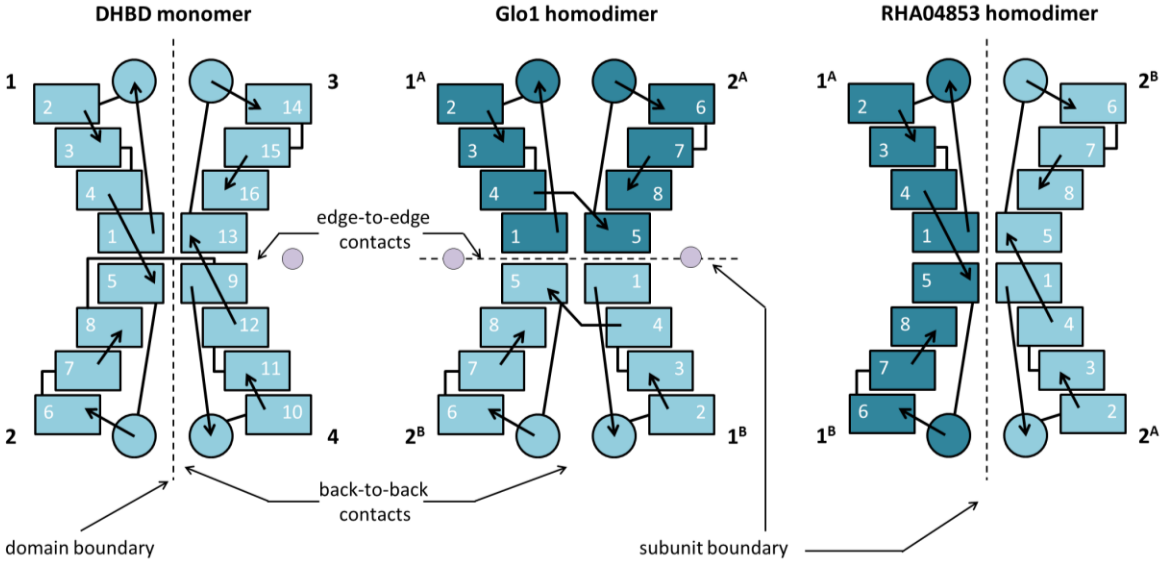


Figure 11: Different dimerization forms of βαβββ-module containing proteins, using DHBH (PDB entry 1HAN), Glo1 (PDB entry 1FRO) and RHA04853 (PDB entry 2U6R) as examples. In this scheme, β-strands are depicted as boxes and α-helices as circles. Connecting segments, shown as lines and arrows, may include additional secondary structure elements. Module numbering is from 1 to 4 in the protein chain, and for Glo1, letters A and B in superscript indicate the monomer. The active sites are marked by violet circles representing the divalent metal ions Fe(II) in DHBD and Zn(II) in Glo1. DHBD and RHA04853 dimerize via back-to-back contacts, while Glo1 dimerizes via edge-to-edge contacts. Figure adapted from Bergdoll *et al.*, 1998.

The second dimerization modus is via edge-to-edge contacts where the α-helices of modules $\mathbf{1^A}$ and $\mathbf{2^A}$ as well as $\mathbf{1^B}$ and $\mathbf{2^B}$ neighboring in sequence are interacting in space (Figure 11, Glo1). However, dimerization by back-to-back contacts is not only possible in βαβββ-module proteins with four modules per monomer (as DHBD), but also in proteins containing two modules per monomer. In this case, the α-helices $\mathbf{1^A}$ and $\mathbf{2^B}$ as well as $\mathbf{1^B}$ and $\mathbf{2^A}$ would interact with each other (Figure 11, RHA04853). Hence, α-helices of modules from different monomers are interacting. The βαβββ-module containing proteins vary in their tertiary and quaternary structure (Figure 12).

Other than dimers, βαβββ-module containing proteins can adapt larger quaternary assemblies which can reach up to octameric complexes of homodimers as observed in DHBD (Han *et al.*, 1995). Some βαβββ-module proteins harbor additional secondary structure elements or have been fused to additional domains due to gene editing. These modifications results in multi-domain proteins with altered functions. One example for such a βαβββ-module protein is the bleomycin-N-acetyltransferase (BAT) from *Streptomyces verticillus* (PDB entry 2ZW6, Oda *et al.*, 2010). This protein combines its βαβββ-module protein domain with an N-acetyltransferase domain, and detoxifies the complex antibiotic bleomycin. Other βαβββ-module proteins harbor metal coordination sites in which divalent transition metal cations are chelated. These cations are essential for the catalytic mechanism, for instance Mn(II), Zn(II) or Fe(II). In DHBD, the active site contains an Fe(II) ion and in human Glo1 each active center coordinates a Zn(II) ion.

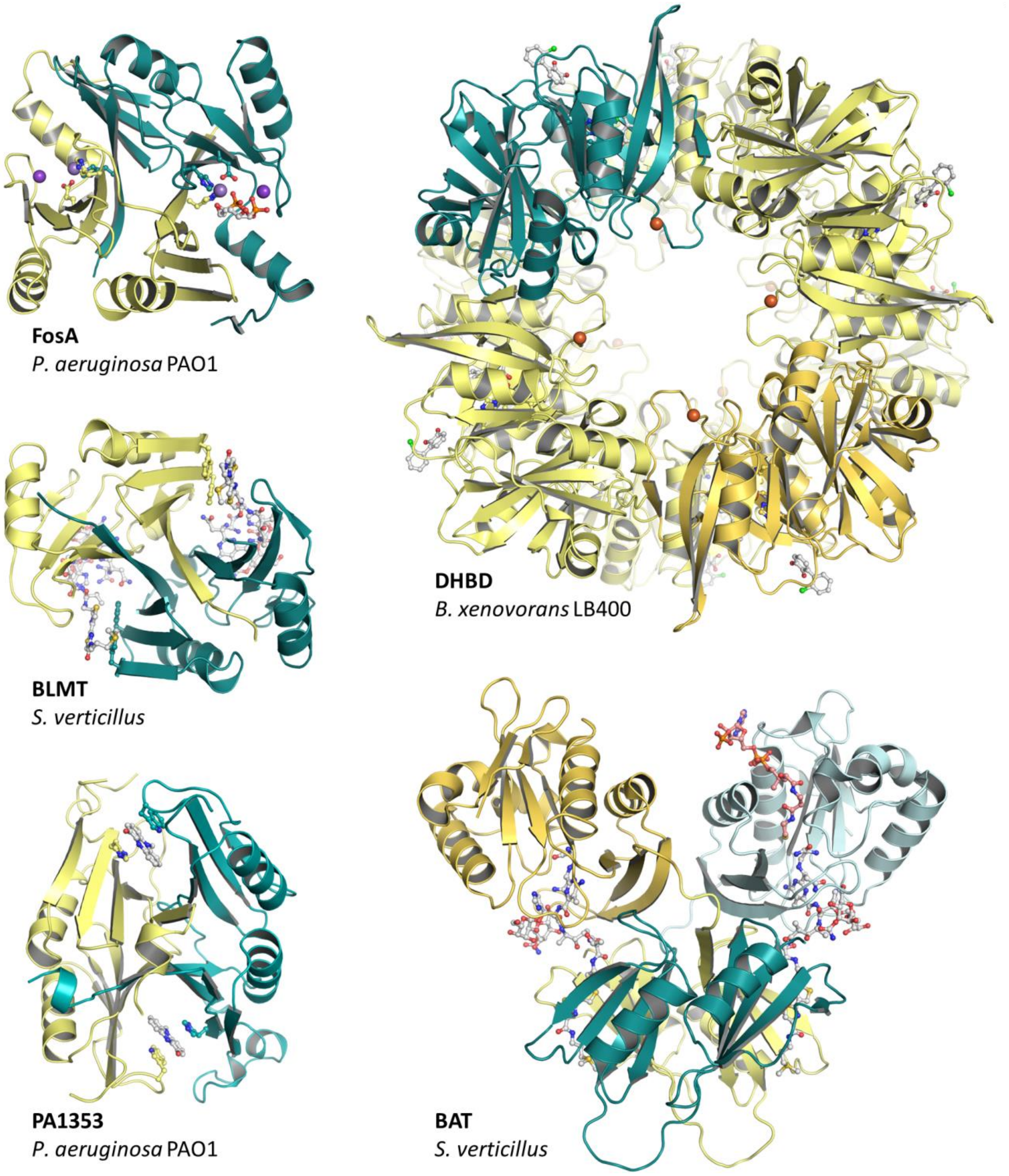


Figure 12: Structural and functional diversity of βαβββ-module containing proteins from *P. aeruginosa* PAO1. FosA from *P. aeruginosa* harbors a catalytic Mn(II) and a K$^+$ ion (violet) and detoxifies fosfomycin. DHBD from *B. xenovorans* LB400 degrades polychlorinated biphenyls using a catalytic Fe(II) and forms an octamer of homodimers. PA1353 is a pyocyanin-binding protein from *P. aeruginosa* which involves two aromatic residues in pyocyanin binding via π-stacking interactions. BLMT from *Streptomyces verticillus* similarly acts as a chaperone for bleomycin. In contrast, BAT from *S. verticillus* acetylates bleomycin under acetyl-CoA consumption, using its additional N-acetyltransferase domain.

Another example for a metal-binding βαβββ-module containing protein is FosA from *P. aeruginosa*, which coordinates a Mn(II) and a K^+ ion to catalyze the addition of GSH to the broad spectrum antibiotic fosfomycin (Rife *et al.*, 2002). In contrast, some βαβββ-module proteins do not harbor metal cations or additional domains, but are involved in binding aromatic compounds, as previously mentioned. Examples for such proteins are PA1353 from *P. aeruginosa*, BLMT, a bleomycin chaperone-like protein from *S. verticillus* (PDB entry 1NIQ, to be published) or MRD, a protein from *Streptomyces lavendulae* involved in mitomycin C binding (Martin *et al.*, 2002). Taken together, βαβββ-module containing proteins adopt a large variety of tertiary and quaternary assemblies which enable proteins of this folding family to execute various tasks and involve in a huge variety of cellular processes, for instance in detoxification, and are therefore important to investigate. Possibly, *P. aeruginosa* possesses a variety of βαβββ-module containing proteins involved in diverse resistances that would be important to discover and characterize.

3 Research Objectives

The opportunistic pathogen *P. aeruginosa* is the causative agent of acute and chronic infections, especially in immunocompromised humans. The bacterium utilizes an arsenal of virulence factors as well as antibiotic resistance factors to facilitate survival and host infection. Most notably, it produces its own antibiotics, the phenazines. Among those, the blue phenazine pigment pyocyanin plays a crucial role in *P. aeruginosa* lung infections. Interestingly, pyocyanin assumes multiple physiological functions, for instance as a signaling molecule in QS, catalyst of redox reactions and regulator of biofilm morphology. Since pyocyanin is also able to cause oxidative stress, *P. aeruginosa* has to shield it intracellularly until it is secreted. Prior to this study, three pyocyanin-binding proteins PA0803, PA1353 and PA4641 from *P. aeruginosa* PAO1 were identified and characterized by X-ray crystallography (Yu, 2009; Kalawy-Fansa, 2010). Still, a crystal structure of PA0803 in complex with pyocyanin in 100% occupancy is missing and will be determined during this study. Furthermore, all three proteins will be tested for binding of aromatic compounds using ITC experiments. Their function as hypothetical pyocyanin resistance proteins will also be tested *in vivo* in *E. coli* and *P. aeruginosa*.

The potential pyocyanin resistance proteins belong to the glyoxalase superfamily or the βαβββ-module containing resistance proteins. In this study, further βαβββ-module containing resistance proteins encoded in the genome of *P. aeruginosa* will be identified and classified. A selection of these proteins will be investigated in order to reveal the resistance mechanisms these proteins are involved in. Using X-ray crystallography, further proteins of this folding family will be structurally elucidated. Some of these proteins will be characterized with using bioinformatics tools as well as biophysical methods such as ITC and microbiological methods to confirm their function *in vivo*.

This study will identify further βαβββ-module containing proteins in *P. aeruginosa* that potentially mediate resistance probably not only against pyocyanin, but against other antibiotic compounds. In the future, this approach will help to understand bacterial resistance mechanisms and possibly lead to novel anti-infective strategies.

4 Principles of Protein X-ray Crystallography

The foundations of X-ray crystallography were provided more than 100 years ago, when Max von Laue carried out the first X-ray diffraction experiment in 1912. In 1914, he was awarded the Nobel Prize in Physics for his discovery of the diffraction of X-rays by crystals. In 1958, the first crystal structure of a protein, sperm whale myoglobin, was determined (Kendrew *et al.*, 1958). Since then, protein X-ray crystallography and structural biology underwent and still undergo huge developments, pushed by the advance of computers and technical equipment. Furthermore, crystallography provides the foundation of modern structural biology techniques and structure-based drug discovery. It can provide detailed molecular insight even into large biological assemblies.

4.1 Crystallization of Proteins

Unfortunately, the process of protein crystallization is empiric and non-deterministic, and it is only possible to increase the probability of a protein to crystallize, but it can never be guaranteed that the process will actually yield protein crystals (Rupp, 2009). Prior to protein crystallization, construct design, cloning and test expressions are fundamental to most of the crystallographic projects. Depending on the protein, this task can be easy or rather challenging. Still it is necessary, since the protein sample has to meet high standards, for instance conformational homogeneity and high purity.

Protein crystallization starts with a sufficient amount of protein solution of suitably high quality and concentration (usually in the mg/mL range). The protein sample is usually mixed in a 1:1 ratio with screening solution, containing a mixture of precipitants, buffers and salts. Then, the protein/precipitant mix is equilibrated against a reservoir of screening solution. The water vapor equilibrates from the protein/precipitant drop to the reservoir containing higher precipitant concentration. Thus, this technique is called the *vapor diffusion method*. It is the most common method to set up protein crystallization trials (Figure 13A). In the vapor diffusion method, the equilibration process reduces the protein solubility in the hope that the protein separates from solution while self-assembling into its crystalline form. The protein solution becomes supersaturated and metastable. Given that the condition is favorable, nucleation sites appear and crystals start to grow. Crystal growth continues until the crystalline material is in equilibrium with the surrounding solution. At this point, the crystal has reached its final size. Nucleation and crystal growth are often described in crystallization diagrams (Figure 13B).

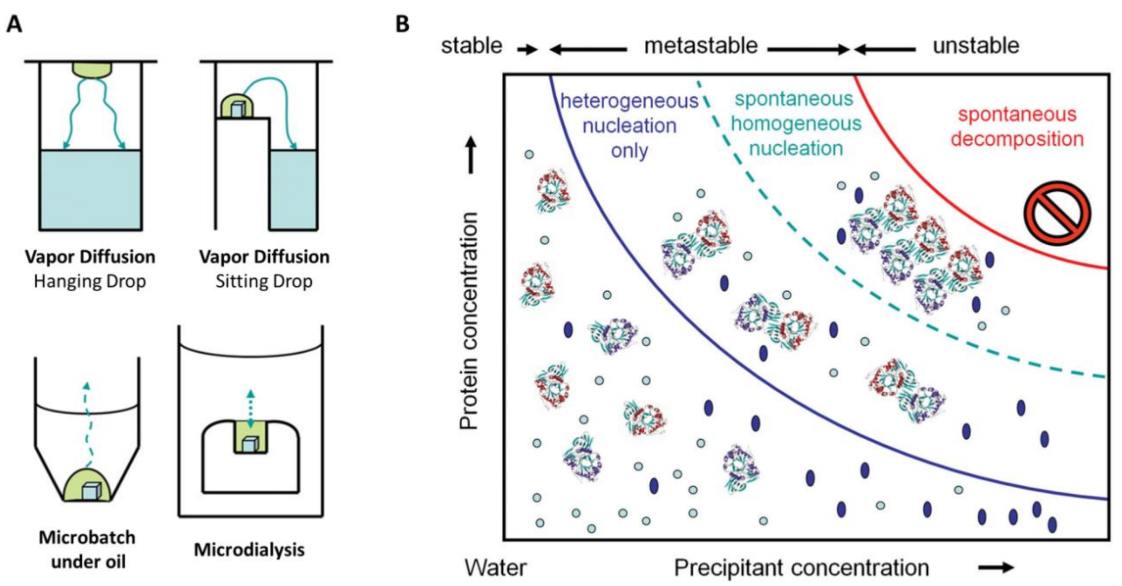


Figure 13: Protein crystallization. A: Schematic views of common crystallization methods. The sitting drop vapor diffusion is commonly used for high-throughput crystal screening, while hanging drop set-ups are used for manual optimization of crystallization experiments. B: Crystallographic phase diagram. Usually, higher supersaturation is needed for homogenous nucleation, while low supersaturation requires external crystallization seeds (for instance microcrystals). If the protein or precipitant concentration is too low, the protein will remain in solution, if it is too high, the protein will precipitate. Figure adapted from Rupp, 2009.

Fortunately, protein crystallization is an automatable process, and nowadays most crystallization screens are set up using robots. The advantages of using robotics are the fast setup of multiple screening conditions at a time, usually in a 96-well plate, and the low amount of protein sample and crystallization solution consumed, in the order of 100 nL of protein solution per drop. The crystallization trials are inspected for crystal formation on a daily basis, often aided by an automated imaging system. Using such imaging systems helps to track and score the appearance of each crystallization drop.

Once a crystallization hit has been identified in a crystallization trial, the crystals often have to be optimized in shape, size and ultimately in diffraction quality. In the optimization plate, the original condition is varied in composition, regarding the pH of the buffer component or concentration of precipitants. In general, components can be added, their concentrations can be varied. Optimization plates can also be used for soaking experiments with ligand or heavy atom complexes. Once a protein crystal has been obtained, it is isolated using a nylon loop. After placing the crystal in a cryoprotecting solution, it is flash cooled in liquid nitrogen. Diffraction experiments are usually performed at cryogenic temperatures. Unfortunately, cooling of protein crystals to these temperatures will lead to the formation of amorphous ice. This ice will in turn lower the diffraction. However, cryogenic temperatures mostly avoid the effects of radiation damage. Radiation damage is caused by radicals that are induced by high energy X-rays. As these radicals can migrate through the crystal, they cause radiation damage which reduces the diffraction quality of the crystals. Taken together, data collection at cryogenic temperatures often prolongues the lifetime of the crystal which in turn has to be protected with cryogenic agents. For instance, these agents can be glycerol or small polyethylene glycols (PEGs).

4.2 Diffraction of X-rays

In order to determine the structure of a protein on the molecular or atomic level, some intrinsic difficulties must first be considered. For instance, the chemical bonds lengths between two atoms are in the range of only a few angstroms Å. The smallest distance between two points in the sample of interest that is able to be distinguished determines the "resolution" of the crystal or protein structure. Among other parameters, the resolution depends on the wavelength of the utilized light. Hence, light of a wavelength of a few angstroms is needed to determine distances of a similar length. Thus, X-rays (λ = 0.1-100 Å) are suitable for protein structure determination on the molecular level. Typically, an X-ray diffraction experiment on a protein crystal is done with X-rays of a wavelength of λ = 1.54 Å (Cu K_α radiation) at an X-ray home sources. Alternatively, synchrotron radiation can be used for such experiments. However, X-rays are only weakly diffracted by organic materials. Only 5% of the X-ray beam are actually scattered by the electrons in matter, while 95% pass through the sample. Hence, it is necessary to raise the signal of the diffracted X-rays to a detectable level. For this purpose, sensitive detectors are needed to collect the diffraction data. Additionally, X-ray crystallography makes use of the physical principle of the positive interference of waves. The X-ray signal is amplified by the positive interference of waves scattered by many molecules of the same type arranged in a regular array. In fact, protein crystals are such periodic arrays of large and flexible macromolecules, held together by weak intermolecular interactions (Rupp, 2009). Hence, they are particularly suited to amplify the signal of diffracted X-rays which can then be detected.

4.3 Bragg's Law, Reciprocal Space and the Ewald Construction

When X-rays hit a protein crystal, they are scattered by the electrons in the protein molecules. These protein molecules can be imagined to lie on crystallographic lattice planes in a protein crystal. The X-rays are then diffracted at these lattice planes (Figure 14).

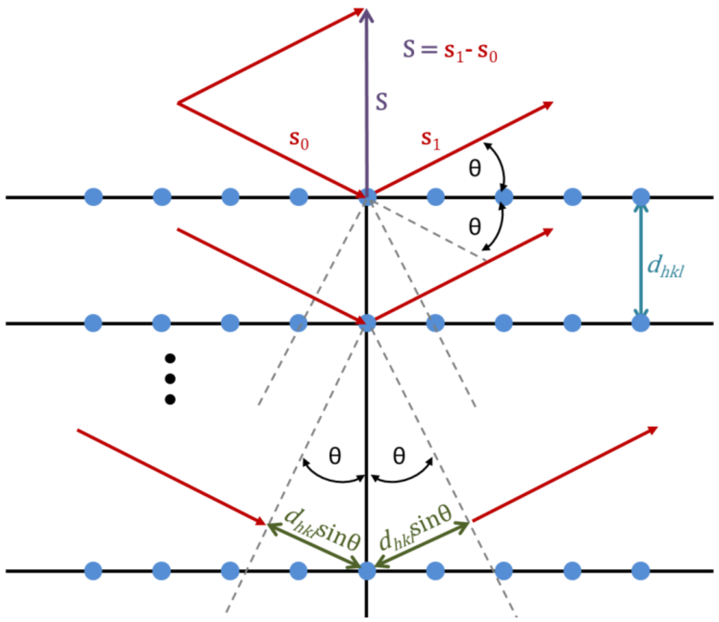


Figure 14: Graphical construction of Bragg's Law. The distance between two crystal planes is d_{hkl}, and θ is the scattering angle. Hence, the total path difference between two scattered X-rays is equal to $2d_{hkl} \sin \theta$ (green). For maximum positive interference, $2d_{hkl} \sin \theta$ has to be equal to a multiple n of the X-ray wavelength λ.

The diffracted X-rays have the same wavelength as the incident beam, but they differ in path length. The interference between X-rays scattered on crystal planes is described the Bragg equation

$$\boldsymbol{n\lambda = 2d_{hkl}\, sin\, \theta} \tag{1}$$

where λ is the wavelength of the X-ray beam, d_{hkl} is the interplanar distance and θ is the scattering angle. The total path difference between two X-ray waves is $2d_{hkl} \sin \theta$, and this term has to be equal to a multiple *n* of the X-ray wavelength λ for maximum positive or *constructive* interference.

When an X-ray beam hits a crystal lattice, some of the scattered waves will interfere constructively and add up to a detectable signal. Most of them, however, will interfere destructively and thus be cancelled out. For this reason, a distinct diffraction pattern can be detected (Figure 15).

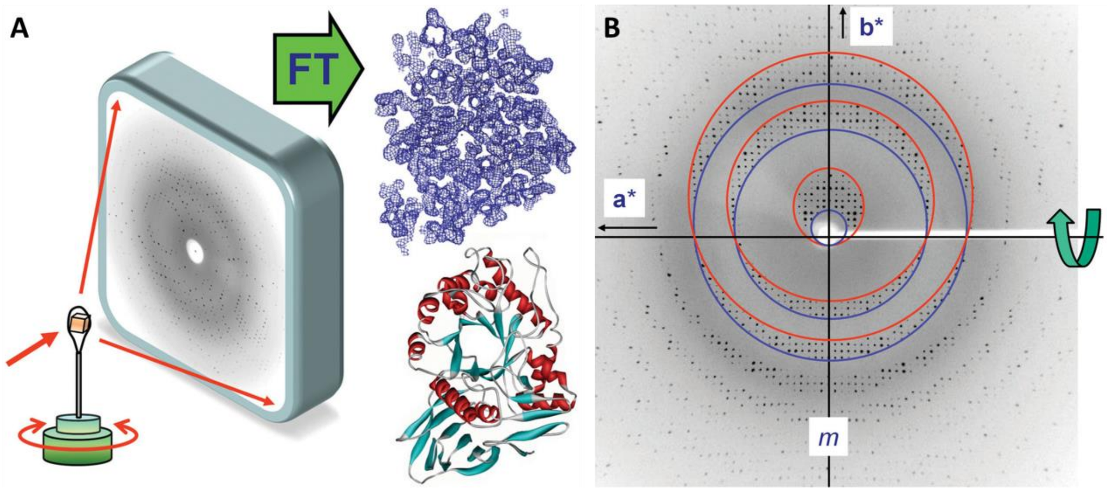


Figure 15: Diffraction of X-rays hitting a protein crystal. A: Only 5% of the incident beam is scattered and can be detected as a distinct diffraction pattern. The pattern is a Fourier transform (FT) of the electron envelope of the crystal content. After a FT of the pattern into the electron envelope, the structural model of the protein can be build. B: Diffraction pattern of a protein crystal. The *a*-axis of the crystal is aligned rotation axis (green arrow) of the crystal. Hence, in this orthogonal crystal, the *c*-axis is aligned to the incident beam. In the diffraction pattern, the mirror symmetry *m* indicates a 2-fold axis in the crystal system. The distances between the reflections are reciprocal to the unit cell axis parameters of the crystal. Figure adapted from Rupp, 2009.

The diffraction pattern is made up of single spots that are called *reflections*. The pattern only contains information about the unit cell parameters and the space group. In contrast, the intensity of the measured reflections in the pattern contains information on the shape of the electron envelope of the protein. However, it is not possible to visualize the protein and its orientation and packing in the crystal without correct analysis and interpretation of the unit cell parameters and space group. Since the relationship between the lattice plane distance d_{hkl} and the spacing of the spots is reciprocal, the description of diffraction events and the following mathematical calculations can be simplified by defining the *reciprocal space*. The reciprocal space is the space that contains the possible spots of interference of the X-ray beam and the crystal lattice. Each direct crystal lattice (in real space) is associated with a reciprocal crystal lattice (in reciprocal space). Hence, the latter can directly be constructed from the real space lattice. A simplified two-dimensional construction of reciprocal space, presented in Figure 16, can also be interpreted as a projection of a three-dimensional lattice down the c-axis.

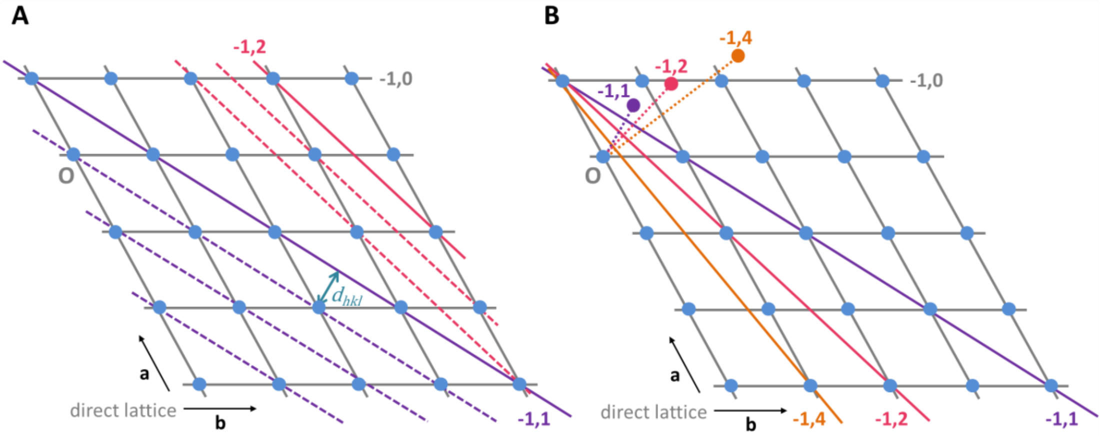


Figure 16: Two-dimensional construction of reciprocal space. The direct lattice is represented in grey with blue lattice points. O is the origin of the direct lattice, the cell axis c is perpendicular to the a/b plane. A: The pink and violet lines each represent a set of direct lattice planes. The distance between two lattice planes of one lattice plane set is d_{hkl}. B: In order to construct reciprocal space, normals to all direct lattice planes are constructed. Each normal terminates at a distance $1/d_{hkl}$ from the origin. The end point of each normal is a reciprocal lattice point. The reciprocal lattice points constitute the reciprocal lattice.

In real space, meaning in a direct crystal lattice, each lattice point can be indexed relative to a chosen origin O, either using direct Weiss indexing (uvw) or reciprocal Miller indexing (hkl). In order to construct the reciprocal lattice, all possible direct lattice planes with Miller indices hkl are constructed. In Miller indexing, index h, k or l is defined as the number of how many times the a-, b- or c-axis of the unit cell is cut by a set of lattice planes, respectively. A set of lattice planes is defined as parallel, equidistant planes with Miller indices hkl. The index h, k or l is zero, if a set of lattice planes is parallel to the respective unit cell axis a, b or c, respectively. After defining all possible lattice planes hkl, normals to these lattice planes are constructed. Each normal terminates at a point of distance $1/\mathbf{d}_{hkl}$ from the origin O where the interplanar distance vector $\mathbf{d}_{hkl}$ is perpendicular distance between two lattice planes of a given set of lattice planes hkl. Such a termination point represents a set of direct lattice planes and is defined as a reciprocal lattice point. The reciprocal lattice points are assigned the indices of the direct planes they represent. Together, the reciprocal lattice points constitute the reciprocal lattice. Having constructed the reciprocal space, it is easier to describe the geometric conditions under which a diffraction event occurs. As described previously, diffraction can only occur if the Bragg equation $n\lambda = 2d_{hkl} \sin \theta$ is satisfied. The equation can be rewritten to

$$\vec{d}^*_{hkl} = 1/\vec{d}_{hkl} = 2sin\theta/n\lambda \quad (2)$$

because the magnitude of each reciprocal lattice vector $\mathbf{d}^*_{hkl}$, connecting the origin O with a reciprocal lattice point hkl, is reciprocal to the interplanar distance vector $\mathbf{d}_{hkl}$. The nth-order diffraction condition can be interpreted as diffraction originating from an n-times higher order set of lattice planes. Therefore, n can be eliminated from the equation (Rupp, 2009). For $n = 1$, the Bragg equation can be rewritten to

$$\vec{d}^*_{hkl} = 2sin\theta/\lambda \quad (3)$$

This form of the Bragg equation can easily be visualized geometrically by the Ewald construction (simplified in Figure 17).

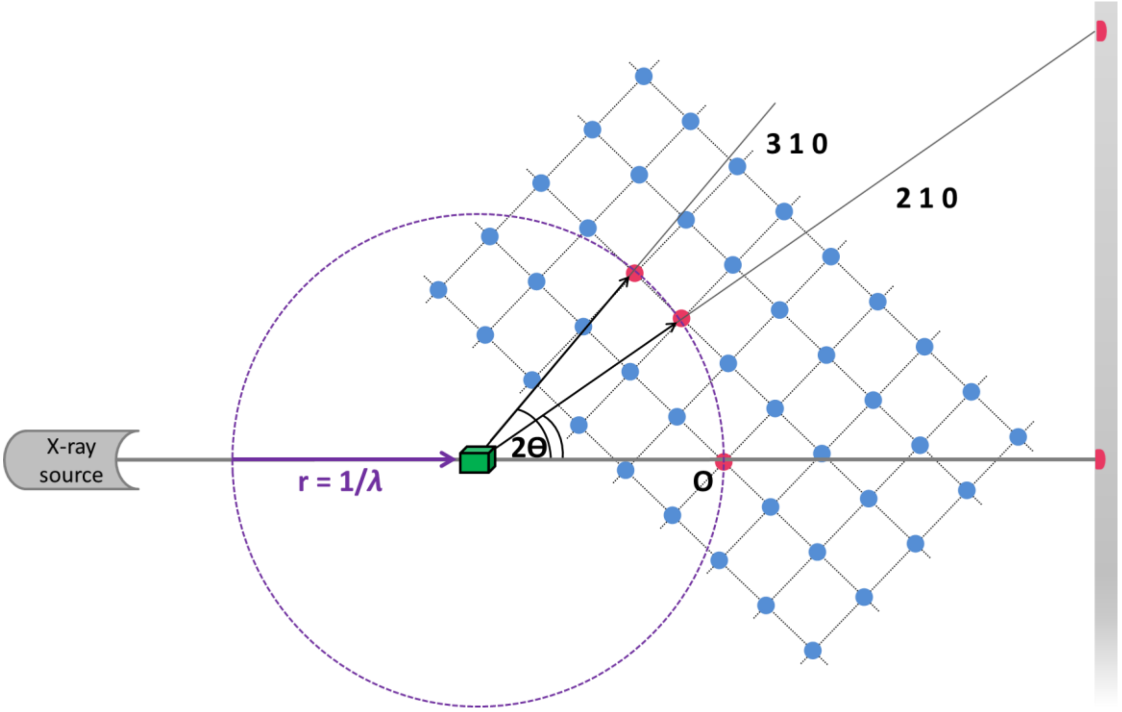


Figure 17: The Ewald construction. The incident beam (grey) is passing through the crystal (green). The crystal's reciprocal lattice is shown in black with blue lattice points. Around the crystal, the Ewald's sphere is constructed with a radius $r = 1/\lambda$, passing through the origin O. If a reciprocal lattice point hkl (for instance 210) coincides with the Ewald's sphere, the Bragg equation is satisfied and constructive interference is observed under the scattering angle 2θ. This means, that a reflection hkl can be observed on the detector, which represents the scattering from the set of direct crystal planes.

In order to describe a condition under which diffraction occurs, a simplified two-dimensional Ewald construction is used (Figure 17). First, the crystal is represented by its reciprocal lattice. An origin O is chosen in the reciprocal lattice. The origin O is hit by the incident beam. An Ewald’s circle (sphere in 3D) with the radius $\mathrm{r} = 1/\lambda$, centered on the X-ray beam, passes through the origin 0. If a reciprocal lattice point coincides with the Ewald’s circle (sphere), for instance 210 or 310 in Figure 17, the condition of the Bragg equation is satisfied and constructive interference is observed under the scattering angle 2θ. Hence, a reflection is detected for these reciprocal lattice points, each representing a set of direct lattice planes. The intensity of the observed reflection depends on the electron distribution on the respective set of lattice planes.

The protein crystal is rotated during X-ray data collection. Hence, the Ewald’s sphere can potentially intersect with each reciprocal lattice point hkl with, thereby producing a reflection hkl on the detector. As the diffraction experiment proceeds, theoretically every possible reciprocal lattice point can produce a reflection on the X-ray detector. However, the maximum resolution of a crystal at a given wavelength λ is defined if the Bragg equation is rewritten to

$$\boldsymbol{d_{min} = n\lambda/2sin\theta_{max}} \tag{4}$$

The maximum of diffraction angle θ_{max} is 90°, hence the maximum of $\sin\theta$ is 1. For $n = 1$, the Bragg equation can be written as

$$\boldsymbol{d_{min} = \lambda/2} \tag{5}$$

This equation explains why X-rays (λ = 0.1 - 100 Å) are required to visualize the molecular detail of proteins (bond lengths are usually in the range of a few angstroms). The crystal’s resolution, meaning the distance between two points in a crystal that can be distinguished, can alternatively be described as determined by the distance d_{hkl} between two lattice planes belonging to the same set of lattice planes. This means, that the resolution for each reflection hkl is determined by the corresponding interplanar distance d_{hkl}. The higher the reflection indices hkl, the smaller is the distance $\mathrm{d_{hkl}}$ and thus, the finer is the structural detail that can be observed.

4.4 The Crystallographic Phase Problem and Its Solution

4.4.1 The Phase Problem

Protein structure determination using crystallography relies on X-ray “light”. Light can be described as a wave, using the wave vector $\mathbf{s_0}$. Waves have a periodic character, and can thus be described as complex numbers (Figure 18). This description helps to understand a central problem in protein crystallography, the *crystallographic phase problem*. If waves are described as complex numbers, they are defined by an amplitude |**F**| and a phase angle φ.

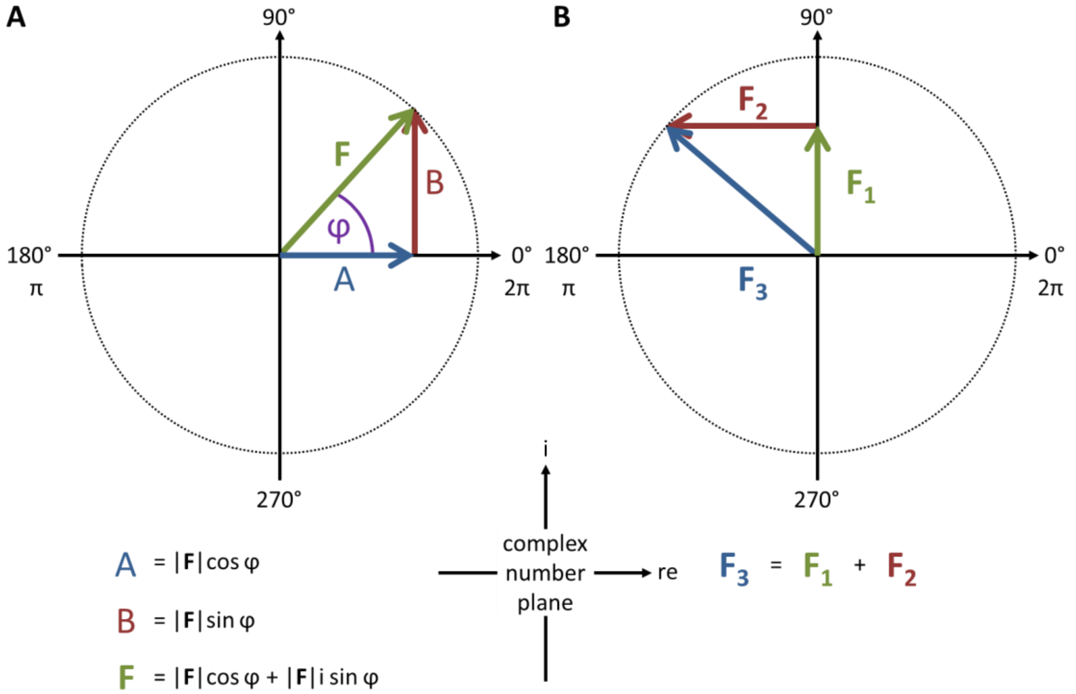


Figure 18: X-ray waves in the complex number plane. A: In the complex number plane, an X-ray wave is described by the vector F. F is formed by a real part A and an imaginary component B. The length of F is given by the amplitude |F| and its direction is given by the phase angle φ. B: The interference of two X-ray waves described by vectors F_1 and F_2 results in a wave described by vector F_3.

In the complex number plane, an X-ray wave is described by the vector **F** with a length of |**F**| and a phase angle φ (Figure 18A). The vector **F** is the sum of the real part A and the imaginary component B (the latter multiplied with the imaginary number i). Using trigonometry, the real part A of the vector **F** is equal to |**F**|cos φ and the imaginary component B is equal to |**F**|sin φ. Hence, **F** can be written as

$$\vec{F} = |\vec{F}| \cos \varphi + |\vec{F}| i \ \sin \varphi \qquad (6)$$

Additionally, the complex number plane can be used to describe the interference of two X-ray waves (Figure 18B). The vectors $\mathbf{F}_1$ and $\mathbf{F}_2$ describing these two X-ray waves add up to a new vector $\mathbf{F}_3$. For instance, X-ray waves are scattered on different electrons of the same atom. The scattered waves have to be added up to define the entire scattering of the electron shell of this atom. This addition results in the atomic scattering factor f_S

which is a result of the integration over the entire volume of the atom. Thus, the atomic scattering factor f_S is defined by the function F_S as

$$\boldsymbol{F_S = f_S = \int_{\vec{r}}^{V} \rho(r) e^{2\pi i \vec{S}\vec{r}}\, d\vec{r}} \tag{7}$$

when V is the volume of the atom, $\mathbf{r}$ is any given position in the atom, $\rho(\mathbf{r})$ is the electron density, and $\mathbf{S}$ is the difference between the vectors $\mathbf{s}_0$ (describing incident X-ray wave) and $\mathbf{s}_1$ (describing the scattered X-ray wave). However, a protein consists of many atoms, and the unit cell of a protein crystal may be formed by more than one protein molecule. Thus, the waves scattered by all electrons in the molecule add up to

$$\boldsymbol{\vec{F}_{\vec{S}} = \sum_j f_{\vec{S},j} e^{2\pi i \vec{S}\vec{r}_j}} \tag{8}$$

for an entire unit cell, when j is the atoms contributing to the intensity of the scattered wave, $f_{S,j}$ is the scattering factor of an atom j, and $2\pi\mathbf{S}\mathbf{r}_j$ is the relative phase each contribution to the scattering factor, with $\mathbf{r}_j$ as the position vectors of all atoms. This equation can be reformulated for reciprocal space using the reciprocal Miller indices hkl (as vector $\mathbf{h}$) for substitution of the scattering vector $\mathbf{S}$ as well as fractional coordinates xyz (as vector $\mathbf{x}$) for substitution of the position vector $\mathbf{r}$. This substitution results in the structure factor $\mathbf{F}_h$ expressed as a function of the index vector $\mathbf{h}$

$$\boldsymbol{\vec{F}_{\vec{h}} = \sum_j f_{\vec{S},j} e^{2\pi i \vec{h}\vec{x}_j}} \tag{9}$$

of a reflection $\mathbf{h}$. In order to include the entire content of the unit cell, it is written as integration over the whole unit cell volume

$$\boldsymbol{\vec{F}_{\vec{h}} = V \int_{x=0}^{1} \rho(\vec{x})\ e^{2\pi i \vec{h}\vec{x}_j} d\vec{x}} \tag{10}$$

The structure factor $\mathbf{F}_h$ is a Fourier transform of the electron density $\rho(\mathbf{x})$. If all structure factors $\mathbf{F}_h$ are Fourier-transformed together, this results in the electron density

$$\boldsymbol{\rho(\vec{x}) = \frac{1}{V} \sum_h \sum_k \sum_l \vec{F}_{\vec{h}} e^{-2\pi i (hx+ky+lz)}} \tag{11}$$

However, as the X-ray wave, a structure factor $\mathbf{F}_h$ is a complex number. When X-ray waves hit a protein crystal, the intensities I_h of the X-ray waves are recorded on a detector as reflections hkl. The intensity I_h of each reflection hkl is equal to the square product of the respective structure factor $\mathbf{F}_h$, meaning

$$\boldsymbol{I_{\vec{h}} = \vec{F}_{\vec{h}} \vec{F}^*_{\vec{h}}} \tag{12}$$

using a simplified description or, because $\mathbf{F}_{hkl}$ is a complex number,

$$\boldsymbol{I_{\vec{h}} = |\vec{F}_{\vec{h}}| e^{i\alpha_{hkl}} |\vec{F}^*_{\vec{h}}| e^{-i\alpha_{hkl}}} \tag{13}$$

where $|\mathbf{F}_h|$ is the amplitude (the real part) and $i\alpha_h$ is the phase (imaginary component) of the structure factor $\mathbf{F}_h$. Complex conjugation of the vectors and subsequent multiplication eliminate the phase information from the intensity I_h. Thus, the I_h only contains information on the structure factor amplitude $|\mathbf{F}_h|$. This *crystallographic phase problem* is also explained by the centrosymmetry of the diffraction pattern.

4.4.2 Diffraction Data Quality

The correct interpretation of the electron map is crucial to obtain a reasonable structural model. However, the quality of the collected X-ray diffraction data is a deciding factor in protein structure determination. For this purpose, data quality is already monitored during data collection using mathematical parameters and crystallographic R-factors. Among the mathematical parameters, the signal to noise ratio $(I/\sigma(I))$, the completeness of the data (every unique reflection hkl has been collected) and the multiplicity or redundancy N (how often every unique reflection hkl has been recorded) are essential to define a resolution cutoff and assess the data quality. For instance, the R-factor

$$\boldsymbol{R} = \sum_{hkl}\sum_{j}\left|\boldsymbol{I}_{hkl,j} - \langle \boldsymbol{I}_{hkl}\rangle\right| / \sum_{hkl}\sum_{j}\boldsymbol{I}_{hkl,j} \tag{14}$$

is such a measure of data quality. However, the R-factor increases with increasing redundancy. This increase is contradictory to the fact that higher redundancy should ensure more accurate data and therefore higher data quality, because every reflection is collected several times. Thus, a redundancy-independent version of the R-factor called R_{meas} can be calculated using

$$\boldsymbol{R}_{meas} = \sum_{hkl}\sqrt{\boldsymbol{N}/\boldsymbol{N}-\mathbf{1}}\sum_{j=1}^{N}\left|\boldsymbol{I}_{hkl,j} - \langle \boldsymbol{I}_{hkl}\rangle\right| / \sum_{hkl}\sum_{j}\boldsymbol{I}_{hkl,j} \tag{15}$$

where $<I_{hkl}>$ is the average of symmetry-related observations of a unique reflection. Calculation of R_{meas} however, results in higher, but more realistic, numerical values than R_{merge} (Diederichs & Karplus, 1997).

In recent years, new R-factors have been developed. One of these R-factors is $R_{p.i.m.}$

$$\boldsymbol{R}_{p.i.m.} = \sum_{hkl}\sqrt{\mathbf{1}/\boldsymbol{N}-\mathbf{1}}\sum_{j=1}^{N}\left|\boldsymbol{I}_{hkl,j} - \langle \boldsymbol{I}_{hkl}\rangle\right| / \sum_{hkl}\sum_{j}\boldsymbol{I}_{hkl,j} \tag{16}$$

which is the precision-indicating merging R-factor of averaged intensities (Weiss, 2001). $R_{p.i.m.}$ takes into account the improvement in precision from multiplicity and is a well-established R-factor in modern crystallography.

4.4.3 Crystallographic Phasing Techniques

The *crystallographic phase problem*, meaning the loss of phase information during data collection, has to be overcome to determine the molecular structure of a protein of interest. For this purpose, the phase information has to be acquired otherwise and fed into the electron density calculations. Three fundamental phasing techniques are known, namely *experimental phasing* using anomalous scatterers, *molecular replacement* and *direct phasing*. While the latter is almost impossible to be used for proteins, the first two methods are frequently used in protein crystallography.

4.4.3.1 Molecular Replacement

Most of the protein structures determined are deposited in the Protein Data Bank (PDB, Berman *et al.*, 2000), which has grown to more than 105 000 entries in early 2015. However, the number of distinct protein folds is not growing as fast. This is probably due to the fact that nature only has evolved a limited number of protein folds. The number of folds known to date varies between 1205 (SCOPe Database 2.04, July 2014) and 1375 (CATH Database 4.0, March 2013), although these calculations do not include all entries in the PDB (Sillitoe *et al.*, 2013; Fox *et al.*, 2014). Hence, if the structure of a protein of interest is to be determined, it is likely that a protein of similar fold has already been deposited in the PDB. Proteins are usually classified as similar, if their structures have a root mean square deviation (r.m.s.d.) of less than 1.5 Å or if their sequences are more than 30% identical. If a similar protein has been identified, it may serve as the s*earch model* in molecular replacement (MR). In the process of MR, the phases of the search model are *borrowed* to determine the phase of protein structure of interest. For this purpose, a calculated electron density is transformed into theoretical structure factors and intensities, which are then compared with the experimental data. However, the orientation of the protein molecule within the unit cell is unknown. For this reason, first a rotational search is performed. Then a translational search is used to define the exact position and orientation of the structure in the asymmetric unit. A correct MR solution is identified using mathematical parameters as the correlation coefficient, the R *factor* or the z-*score*. Additionally, the MR solution can be validated by inspection of the crystal packing and the electron density. Finally, the solution is subjected to the model building and refinement process.

If MR using a single, unmodified MR search model is not successful, it is possible to optimize the MR process. The search model can be modified, for instance by side chain deletion, or multiple models can be generated. For this purpose, MR engines such as *BALBES* (Long *et al.*, 2007) or *MrBump* (Keegan & Winn, 2008) have been developed recently. These engines scan the sequence of interest against a PDB-derived database and derive appropriate search models for MR.

4.4.3.2 Isomorphous Replacement and Anomalous Diffraction

If no suitable MR model is available, the phase information for structure determination of the protein of interest has to be acquired otherwise. This can be accomplished by experimental phasing, which is dependent on heavy atoms introduced into the crystal.

Experimental phasing can be achieved using either single or multiple isomorphous replacement (SIR or MIR). Both SIR and MIR are dependent on successful derivatization of a protein crystal with heavy atoms, which may be difficult. The more electrons the heavy atom possesses, the more suitable it is for SIR and MIR techniques. For instance, mercury salts that modify cysteine residues are used for derivatization. Derivatization is not supposed to affect the crystal contacts or unit cell parameters, meaning that the native and derivative crystals are isomorphous, in order to successfully phase the structure later on. The molecular structure of the protein can then be determined by first subtracting the diffraction patterns of the native and the derivative crystal, and then calculating Patterson map from this difference. This approach will yield the position of the heavy atoms. When using SIR, one native and only one derivative crystal are needed, but the phase information will be ambiguous. When using MIR, more than one derivative crystal is needed in addition to the native crystal. In contrast to SIR, experimental phasing with MIR will lead to unambiguous phase information. However, SIR and MIR have largely been replaced by single and multiple anomalous diffraction (SAD and MAD) techniques in the last decades. Originally, SAD and MAD were invented out of necessity (Hendrickson, 1985). Today, SAD is particularly used on selenomethionine-derivatized crystals or for sulfur phasing using the protein’s intrinsic sulfur atoms. Nevertheless, heavy atom derivatives are still used, and in some cases, an intrinsic catalytic ion, for instance Fe(II) or Zn(II), can be used in these phasing techniques. SAD and MAD phasing take advantage of the centrosymmetry of the diffraction pattern. Friedel’s law

$$\boldsymbol{I_{hkl}} = \vec{\boldsymbol{F}}_{\boldsymbol{hkl}}\vec{\boldsymbol{F}}^{*}_{\boldsymbol{hkl}} = \vec{\boldsymbol{F}}_{\boldsymbol{-h-k-l}}\vec{\boldsymbol{F}}^{*}_{\boldsymbol{-h-k-l}} = \boldsymbol{I_{-h-k-l}} \tag{17}$$

states that a reflection hkl has the same intensity I as its centrosymmetrically related reflection -h-k-l, because hkl and -h-k-l are reflections of the same set of lattice planes. These symmetry-related reflections are called Friedel or Bijvoet pairs (Figure 19A). Friedel’s law only holds if the structure factors $\mathbf{F}_S$ are real functions. In contrast, Friedel’s law is false if the diffraction data is collected at an X-ray absorption edge of an anomalous scatterer (Figure 19B).

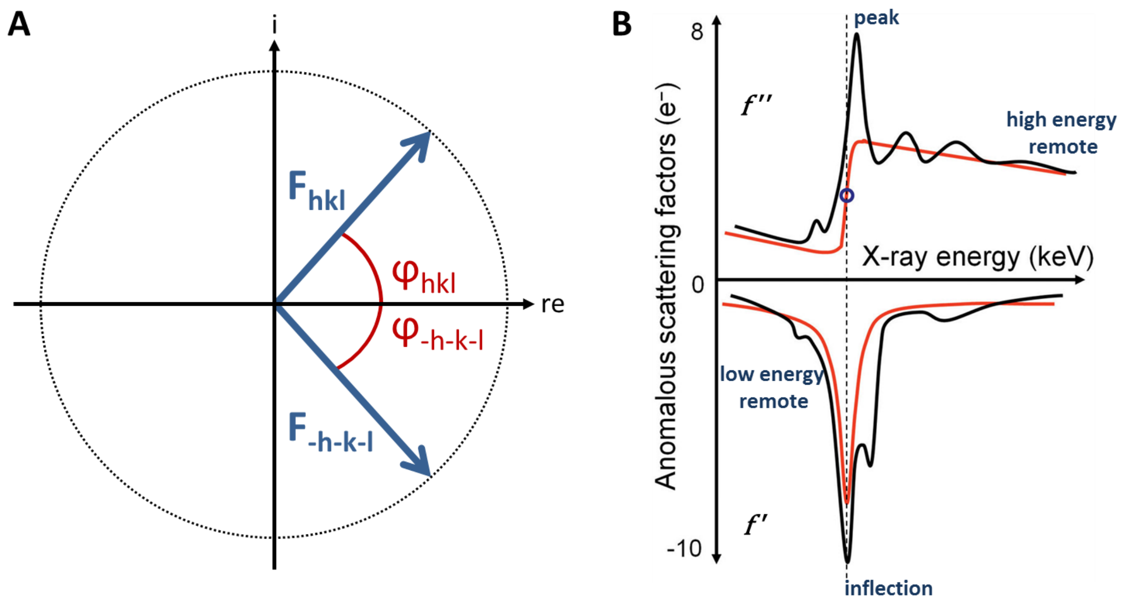


Figure 19: Friedel’s Law. A: Representation of a Friedel pair in the complex number plane. Two reflections hkl and -h-k-l (a Friedel pair) have the same intensity I_{hkl} if their structure factors are real functions. Friedel pairs are reflections originating from the same set of lattice planes, but from mirrored direction. B: An X-ray absorption scan (upper black curve) yields f′′ and can be used to derive f′. The negative peak of the f′ function coincides with the inflection point (blue circle) of the f′′ function. The largest anomalous differences can be obtained from data collected at the absorption peak wavelength. Figure B adapted from (Rupp, 2009)

At X-ray absorption edges of anomalous scatterers, wavelength-dependent changes in the atomic scattering factor f_S occur and can be described as

$$\boldsymbol{f}_{(\vec{\boldsymbol{S}},\boldsymbol{\lambda})} = \boldsymbol{f}^{\boldsymbol{0}}_{(\vec{\boldsymbol{S}})} + \boldsymbol{f'}_{(\boldsymbol{\lambda})} + \boldsymbol{i} \cdot \boldsymbol{f''}_{(\boldsymbol{\lambda})} \tag{18}$$

where $f'_{(\lambda)}$ is the *dispersive* and $f''_{(\lambda)}$ is the *anomalous* contribution to the structure factor. Both $f'_{(\lambda)}$ and $f''_{(\lambda)}$ can be determined experimentally by an X-ray absorption scan, using the tunable wavelength of a synchrotron source (Figure 19B). The phase shift Δφ for $f'_{(\lambda)}$ (Δφ = 90°) breaks down Friedel symmetry and enables the observation of anomalous differences in the dataset (Figure 20A-C).

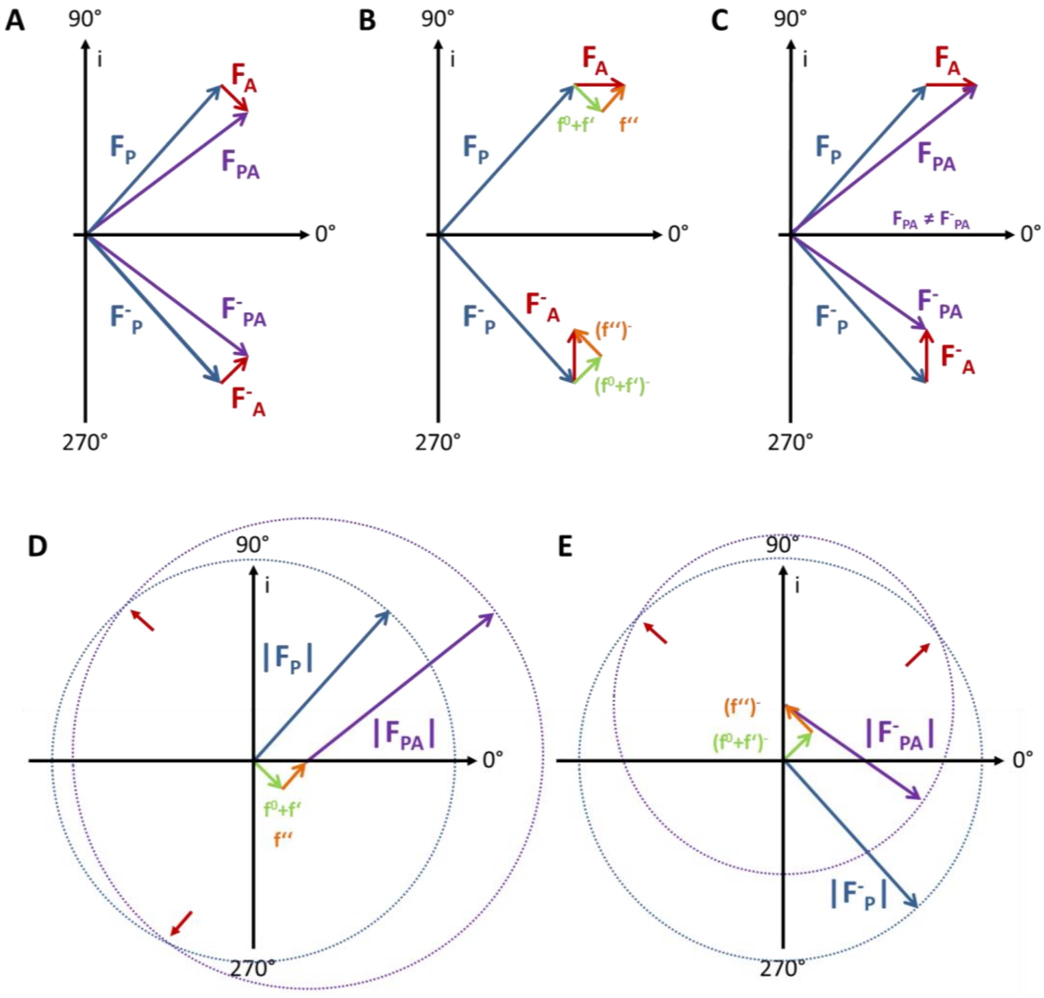


Figure 20: Friedel's law and its breakdown. F_P/F^-_P is the contribution of the protein to the scattering, F_A/F^-_A is the contribution of the anomalous scatterer and F_{PA}/F^-_{PA} is the contribution of both. A: Friedel's law is true if there are no anomalous scatterers. The structure factors F_S are real functions and F_{PA} is equal to F^-_{PA}. B: In the presence of an anomalous scatterer, changes in the atomic scattering factor f_S occur and f' and f'' contribute to the scattering. Because f'' contributes with Δφ = 90°, anomalous differences can be observed in the data. C: When anomalous scattering occurs, F_{PA} is not equal to F^-_{PA} anymore. D & E: Harker constructions yield two solutions for the phase (red arrows) and hence for the position of the anomalous scatterers. One of these solutions can be excludes after refinement because it will not lead to decreasing R-factors.

For SAD phasing, it is crucial to obtain the maximal signal of the anomalous contribution $f''_{(\lambda)}$. For this purpose, e. g. a dataset for selenium SAD should be collected at the peak wavelength, meaning at 0.9795 Å / 12.6578 keV for the selenium K-edge. For MAD phasing, after the peak dataset, the next dataset to be collected is the inflection data set which maximizes the signal of $f'_{(\lambda)}$ (Figure 19B). Finally, a high energy remote dataset is collected, containing information on both $f'_{(\lambda)}$ and $f''_{(\lambda)}$.

The anomalous differences can be used to determine the positions of the anomalous scatterers in the unit cell. The most powerful tool to use is the Patterson function

$$\boldsymbol{P}(\vec{\boldsymbol{u}}) = \boldsymbol{\rho}(\vec{\boldsymbol{r}}) \cdot \boldsymbol{\rho}(-\vec{\boldsymbol{r}}) = \frac{1}{V}\sum_h \sum_k \sum_l |\vec{\boldsymbol{F}}_{hkl}|^2 e^{-2\pi i(hx+ky+lz)} \tag{19}$$

This function can directly be calculated from the measured intensities, meaning that no prior phase information is needed. The Patterson function is centrosymmetric because it describes the autocorrelation of the electron density convoluted with its inverse. The function shows vectors between atoms in the unit cell. However, if there are N atoms inside the unit cell, there are N*(N-1) vectors between them. As only the positions of the heavy atoms should be considered in the Patterson map, the problem is simplified by calculating a difference Patterson map using

$$\Delta\boldsymbol{P}(\vec{\boldsymbol{u}}) = \frac{1}{V}\sum_h \sum_k \sum_l (|\Delta\vec{\boldsymbol{F}}_{PA}| - |\Delta\vec{\boldsymbol{F}}_{P}|)^2 e^{-2\pi i(hx+ky+lz)} \tag{20}$$

with the intensities obtained from the native dataset, and subtracting them from the derivative dataset. By doing so, only the anomalous scatterers are taken into account and there are only few peaks in the Patterson map. Using these peaks, the relative positions of the anomalous scatterers in the unit cell can be calculated.

4.5 Model Building and Refinement

Once the electron density has been obtained by experimental phasing or MR, it can be visualized by model building programs. These programs allow the interpretation of the electron density map. Thus, a structural model can be placed into the electron envelope. In this process, knowledge about proteins in general as well as about the protein of interest is applied to build the structural model. This knowledge includes the protein sequence, the geometry of single amino acids (e. g. proteins only consist of L-amino acids) as well as bonds between them (e. g. all peptide bonds are planar). The unique shape of the electron envelope may help to identify bulky amino acids like tryptophan or secondary structure elements, which underlie distinct geometry restraints. Similar to helices and β-sheet, loops adopt energetically favorable conformations, and the amino acids in the protein are usually present as their preferred rotamers. In recent years, the initial rounds of manual model building are being replaced by automated model building, aided by computer programs. However, automated model building cannot completely replace manual model building done by a scientist. The reason for this is that such programs only work successfully under certain conditions, for instance a resolution of at least 2.7 Å or higher and good-quality initial phases.

In the next step, computer-assisted refinement calculations optimize the geometry of the model, temperature factors and occupancies. For this purpose, a theoretically expected set of structure factors F_{calc} is calculated from the model. Then, the structural model is refined against the experimentally observed structure factors F_{obs}. In the process, the structure factor amplitudes $|F_{calc}|$ and $|F_{obs}|$ are constantly compared and their differences are minimized, guiding the structure towards a global minimum. Additionally, the refinement progress is monitored by the R-factor

$$\boldsymbol{R} = \sum_{\boldsymbol{hkl}} ||\vec{\boldsymbol{F}}_{\boldsymbol{obs}}| - \boldsymbol{k}|\vec{\boldsymbol{F}}_{\boldsymbol{calc}}|| / \sum_{\boldsymbol{hkl}} |\vec{\boldsymbol{F}}_{\boldsymbol{obs}}| \tag{21}$$

where k is a scaling factor. In order ensure correct structure optimization, 5% of the reflections are selected for the R_{free} calculations (equation 21) and thus not included into R-factor calculations. Only if both R-factor and R_{free} decrease, the structure refinement is successful. In some cases, the refinement process is additionally facilitated by density modification or non-crystallographic symmetry (NCS). For instance, if there is more than one copy of a molecule in the asymmetric unit, the copies are related to each other using NCS operators. By defining an NCS mask around the molecule and using the NCS operator, the electron density can then be improved by NCS averaging.

The structural model is optimized in alternating refinement and model building steps. In the next model building step, two types of electron density can be displayed. First, $(2m|F_{obs}|\text{-}D|F_{calc}|)e^{i\alpha_{calc}}$ density is displayed at 1σ. This density surrounds the whole structural model in areas where $|F_{obs}|$ is equal to $|F_{calc}|$. Second, positive $(m|F_{obs}|\text{-}D|F_{calc}|)e^{i\alpha_{calc}}$ electron density shows that additional features need to be added to the model, for instance side chains or ions, while negative $(m|F_{obs}|\text{-}D|F_{calc}|)e^{i\alpha_{calc}}$ density shows parts of the structure that are incorrectly modeled. Thus, these densities aid to build a correct structural model.

5 Materials and Methods

5.1 Chemicals, Solutions and Materials

5.1.1 Chemicals

If not indicated differently, all chemicals were purchased in the highest purity available (p. a., *pro analysis*) from the companies AppliChem (Darmstadt, Germany), Fluka (Seelze, Germany), Merck (Darmstadt, Germany), Qiagen (Hilden, Germany), Roth (Karlsruhe, Germany) and Sigma-Aldrich (Deisenhofen, Germany). Bleomycin was acquired as Bleomedac® from medac GmbH (Wedel, Germany).

5.1.2 Buffers, Media and Stock Solutions

Buffers and stock solutions were prepared in ddH_2O and filter sterilized. Media were autoclaved. Buffers used for protein preparation and ITC experiments were degased. Antibiotics were used at concentrations of 100 µg/mL ampicillin, 34 µg/mL chloramphenicol, 12 µg/mL gentamicin and 20 µg/mL tetracycline. Stock solutions of hydrophobic ligands were prepared in 100% DMSO, stock solutions of hydrophilic ligands were prepared in H_2O or 20 mM Tris/HCl pH 8, 150 mM sodium chloride.

DNA Isolation	
Genomic DNA buffer	0.8 mL of 20% (*w/v*) sucrose, 50 mM Tris/HCl pH 8; 15 µL of 10 mg/mL lysozyme; 15 mg/mL proteinase K; 40 µL 10% (*w/v*) sodium dodecyl sulfate (SDS)
Agarose Gel Electrophoresis	
50x TAE buffer	242 g Tris, 57.1 mL glacial acetic acid, 100 mL 0.5 M EDTA pH 8 in 1 L ddH_2O
1% agarose solution	1 g agarose in 100 mL of 1x TAE buffer
DNA Standard	100 bp DNA Ladder Plus, Peqlab 1 kb DNA Ladder, New England Biolabs Gene Ruler DNA Ladder Mix, Thermo Scientific

DNA sample buffer	40% glycerol, 0,25% bromophenol blue

Media

LB agar (pH 7,4)	10 g/L NaCl, 10 g/L bacto tryptone, 5 g/L yeast extract, 15 g/L agar
LB medium (pH 7,4)	10 g/L NaCl, 10 g/L bacto tryptone, 5 g/L yeast extract
5x M9 salts	64 g/L Na_2HPO_4 x 7 H_2O, 15 g/L KH_2PO_4, 5 g/L NH_4Cl, 2.5 g/L NaCl in 1 L ddH_2O
Glucose-salt stock	10 mL 20% (*w/v*) glucose, 2 mL 1 M $MgSO_4$, 50 µL 1 M $CaCl_2$
M9-Agar	63,5 ml 2x agar (30 g/L), 25 mL 5x M9 salts, 34.5 mL ddH_2O, 3 mL glucose-salt stock

Purification

Buffer A	300 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4 pH 8
Buffer B	500 mM imidazole in buffer A
Dialysis buffer	150 mM NaCl, 20 mM Tris/HCl pH 8
TEV cleavage buffer	dialysis buffer with 2 mM β-mercapto-ethanol (βME)
SEC buffer	150 mM NaCl, 20 mM Tris/HCl pH 8

SDS-PAGE

4x stacking gel buffer	0.5 M Tris/HCl, 0.4% SDS, pH 6.8
Stacking gel (5%)	5.8 mL ddH_2O, 2.5 ml 4x stacking gel buffer, 1.7 mL 30% (*w/v*) acrylamide (37.5:1), 100 µL 10% (*w/v*) APS, 10 µL TEMED
4x resolving gel buffer	1.5 M Tris/HCl pH 8.8, 0.4% (*w/v*) SDS
Resolving gel (16%)	3.9 mL H_2O, 4.5 mL 4x resolving gel buffer, 9.6 mL 30% (*w/v*) acrylamide (37.5:1), 200 µL 10% APS, 20 µL TEMED
Resolving gel (18%)	2.6 mL H_2O, 4.5 mL 4x separating gel buffer, 10.8 mL 30% (*w/v*) acrylamide (37.5:1), 200 µL 10% (*w/v*) APS, 20 µL TEMED
5x SDS-PAGE sample buffer	50 mM Tris/HCl pH 6.8, 500 mM βME, 50% (*v/v*) glycerol, 10% (*w/v*) SDS, 0.05% (*w/v*) bromophenol blue
Running buffer	25 mM Tris/HCl, 200 mM glycine, 0.1 % (*w/v*) SDS

Staining solution	0.2% (*w/v*) Coomassie Brilliant Blue, 30% (*v/v*) methanol, 10% (*w/v*) acetic acid
Destaining solution	20% (*v/v*) methanol, 10% (*v/v*) glacial acetic acid
Protein Standard	Protein Marker I, 14.4-116 kDa, Peqlab Unstained Protein Molecular Weight Marker, Fermentas

5.1.3 Oligonucleotides

Oligonucleotides used in this study (Table 1) were designed using OligoCalc (Kibbe, 2007) and ordered from MWG Eurofins (Munich, Germany). Melting temperatures were calculated using the *nearest neighbor* method. The ORFs of the proteins were amplified using these oligonucleotides either on *P. aeruginosa* genomic DNA or on plasmid DNA containing the ORF. Sequences of the ORFs were taken from the *Pseudomonas Genome Database* (Winsor *et al.*, 2011).

Table 1: Oligonucleotides used for cloning and sequencing. Endonuclease restriction sites are listed in bold, stop codons are underlined.

Oligonucleotide	Sequence 5'→ 3'	Target vector
PA0803_NcoI_s	AAA TTC **CCA TGG** ACA CGC CCG CCA GCA ACA	pHERD30T
PA0803_HindIII_as	ACG GTC **AAG CTT** TCA TGC CCA GGG ATC ATA G	pHERD30T
PA0880_NdeI_s	ATA GCT **CAT ATG** CAG ATC GAC CGT CTC GAC	pET19m
PA0880_BamHI_as	ATT AAA **GGA TCC** TTA CAG CGG GTT GCT CAG TTC	pET19m
PA1353_NdeI_s	ACC GTC **CAT ATG** CAA ATC GTC CCC TAC CTG ATC TT	pET19m
PA1353_XhoI_as	ATT TTA **CTC GAG** CTA CCG GTC CTG CTC CTC GCA	pET19m
PA1353_NcoI_s	ACC GTC **CCA TGG** AAA TCG TCC CCT ACC TGA TCT T	pHERD30T
PA1353_HindIII_as	ATT CTA **AAG CTT** CTA CCG GTC CTG CTC GCA	pHERD30T
PA1354_NdeI_s	ATC CCA **CAT ATG** AGA TTC ATG GTG ATC GTC AAG GC	pET19m
PA1354_XhoI_as	TCT CTT **CTC GAG** TCA GTT GCG TTT TTC CAG CTC	pET19m
PA3127_NcoI_s	AAT ATC **CCA TGG** TGG AAC GCG CAA TCC CGA	pHERD30T
PA3127_HindIII_as	AA GTC **AAG CTT** TCA GTA GCG AGT GAA GGT CA	pHERD30T
PA3128_XhoI_as	TAT AAT **CTC GAG** CTA TCG TCC TCC GCT GAC ATC	pET19m
PA4183opt_NdeI_16s	AGC TGT **CAT ATG** TCC CAA CCG TTT CTG TGG CGT	pET19m
PA4183opt_NdeI_32s	CCT ATC **CAT ATG** GCC GAA CCG CGT GTG TTC ATT C	pET19m
PA4183opt_XhoI_as	TGG CAA **CTC GAG** TTA TGC TTC GCC CGG TTG C	pET19m
PA4518opt_NdeI_s	AAT GTC **CAT ATG** ACC GGT AAT GGT GTT AGC CTG GT	pET19m

Oligonucleotide	Sequence 5’→ 3’	Target vector
PA4518opt_XhoI_as	AAT GTC **CTC GAG** ATT CGG CAG CAC CAG ATT	pET19m
PA4518opt_QC_s	AAT CTG GTG CTG CCG <u>TAA</u> **CTC GAG** GGA TCC GGC	pET19m_PA4518
PA4518opt_QC_as	GCC GGA TCC **CTC GAG** <u>TTA</u> CGG CAG CAC CAG ATT	pET19m_PA4518
PA4641_NdeI_s	ATA GCT **CAT ATG** TAC GTC CAG CCC TAC GTC TGC	pET19m
PA4641_XhoI_as	AAT TTT **CTC GAG** <u>TCA</u> TGC TCC TGC CTC GCA	pET19m
PA4641_PciI_s	ATA GCA **ACA TGT** ACG TCC AGC CCT ACG TCT G	pHERD30T
PA4641_HindIII_as	AAT GTT **AAG CTT** <u>TCA</u> TGC TCC TGC CTC GCA	pHERD30T
T7_s	TTAATACGACTCACTATAGGGGAA	pET19m
T7_as	CTAGTTATTGCTCAGCGGTGG	pET19m
pHERD_SF	ATC GCA ACT CTC TAC TGT TTC T	pHERD30T
pHERD_SR	TGC AAG GCG ATT AAG TTG GGT	pHERD30T

5.1.4 Plasmids and Constructs

All plasmids used and constructs used in this study are listed in Table 2. Constructs cloned into pET19m or pOPINF are produced with an N-terminal His$_6$-tag and a tobacco etch virus (TEV) protease cleavage site if not indicated differently.

Table 2: Vectors constructed and used in this study. The vector pET19m_PA3127_PA3128 contains the dicystronic operon, amplified from *P. aeruginosa* PAO1 genomic DNA. DPF = Dortmund Protein Facility, MPI = Max Planck Institute.

Plasmid	Purpose	Host	Source
pET19m	Cloning	*E. coli*	MPI Dortmund
pET19m_PA0803	Cloning, Protein Production, Spot Plates	*E. coli*	Yu, 2009
pET19m_PA0880	Protein Production	*E. coli*	this study
pET19m_PA1353	Protein Production, Spot Plates	*E. coli*	this study
pET19m_PA1354	Protein Production	*E. coli*	this study
pET19m_PA1672	Protein Production	*E. coli*	Weber, 2011
pET19m_PA3127	Protein Production, Spot Plates	*E. coli*	Bittner, 2012
pET19m_PA3127_1-114	Protein Production, Spot Plates	*E. coli*	Popp, 2013
pET19m_PA3127_123-267	Protein Production, Spot Plates	*E. coli*	Popp, 2013
pET19m_PA3127_PA3128	Spot Plates	*E. coli*	this study
pET19m_PA4183opt	Protein Production	*E. coli*	this study
pET19m_PA4183opt_16-155	Protein Production	*E. coli*	this study
pET19m_PA4183opt_32-155	Protein Production	*E. coli*	this study

Plasmid	Purpose	Host	Source
pET19m_PA4518	Protein Production	*E. coli*	this study
pET19m_PA4641	Protein Production, Spot Plates	*E. coli*	this study
pHERD30T	Cloning, Spot Plates	*P. aeruginosa*	Qui *et al.*, 2008
pHERD30T_PA0803	Spot Plates	*P. aeruginosa*	this study
pHERD30T_PA1353	Spot Plates	*P. aeruginosa*	this study
pHERD30T_PA3127	Spot Plates	*P. aeruginosa*	this study
pHERD30T_PA4641	Spot Plates	*P. aeruginosa*	this study
pOPINF_PA1353	Cloning	*E. coli*	DPF, MPI Dortmund
pOPINF_PA4641	Cloning	*E. coli*	DPF, MPI Dortmund
pMA-T_PA4518opt	Cloning	*E. coli*	GeneArt, Regensburg
pUC57_PA4183opt	Cloning	*E. coli*	GeneScript, Piscataway, NJ, U.S.A.

Further plasmids constructed in this study can be found in Table 45 in the appendix.

5.1.5 Microorganisms

All microorganisms and generated strains used for the experiments conducted in this study are listed in Table 3.

Table 3: Microorganisms used in this study. *P. aeruginosa* mutants are transposon insertion mutants available from a PAO1 or a PA14 transposon library (Liberati *et al.*, 2006; Held *et al.*, 2012).

Bacterial Strain	Antibiotic Resistance	Purpose	Source
E. coli DH5α	-	Cloning	Invitrogen
E.coli TOP10	Streptomycin	Cloning	Invitrogen
E.coli XL1-Blue	Tetracycline	Cloning	Stratagene
E. coli BL21(DE3)	-	Protein Production	Novagen
E. coli BL21 (DE3) Codon+	Chloramphenicol	Protein Production	Novagen
E. coli BL21 (DE3) pLysS	Chloramphenicol	Protein Production	Novagen
E. coli Rosetta 2	Chloramphenicol	Protein Production	Novagen
P. aeruginosa PAO1 wild type	-	Spot Plate Assay	Collaborator
P. aeruginosa PAO1 pHERD30T	Gentamicin	Spot Plate Assay	this study
P. aeruginosa PAO1 pHERD30T_PA0803	Gentamicin	Spot Plate Assay	this study
P. aeruginosa PAO1 pHERD30T_PA1353	Gentamicin	Spot Plate Assay	this study
P. aeruginosa PAO1 pHERD30T_PA3127	Gentamicin	Spot Plate Assay	this study

Bacterial Strain	Antibiotic Resistance	Purpose	Source
P. aeruginosa PAO1 pHERD30T_PA4641	Gentamicin	Spot Plate Assay	this study
P. aeruginosa PAO1 Δ1353 ID 13110	Tetracycline	Spot Plate Assay	Held *et al.*, 2012
P. aeruginosa PAO1 Δ1353 ID 48549	Tetracycline	Spot Plate Assay	Held *et al.*, 2012
P. aeruginosa PAO1 Δ4641 ID 11551	Tetracycline	Spot Plate Assay	Held *et al.*, 2012
P. aeruginosa PAO1 Δ4641 ID 48777	Tetracycline	Spot Plate Assay	Held *et al.*, 2012
P. aeruginosa PA14 wild type	-	Spot Plate Assay	Collaborator
P. aeruginosa PA14 Δ0803 ID 46327	Gentamicin	Spot Plate Assay	Liberati *et al.*, 2006
P. aeruginosa PA14 Δ3127 ID 05812	Gentamicin	Spot Plate Assay	Liberati *et al.*, 2006
P. aeruginosa PA14 Δ3127 ID 25570	Gentamicin	Spot Plate Assay	Liberati *et al.*, 2006
P. aeruginosa PA14 Δ3127 ID 37013	Gentamicin	Spot Plate Assay	Liberati *et al.*, 2006

5.1.6 Enzymes

All restriction endonucleases (BamHI, HindIII, NcoI, NdeI, PciI, XhoI), *Phusion* DNA Polymerase, *Taq* Polymerase, Antarctic Phosphatase as well as T4 DNA Ligase were ordered from New England Biolabs (Ipswich, MA, U.S.A.). *KAPAHiFi* Polymerase was purchased from Peqlab (Erlangen, Germany). TEV protease was produced in the laboratory from plasmid pHR-MBP_TEV (DPF, MPI Dortmund) as His$_6$-MBP-TEV and purified via Ni(II)-immobilized metal ion affinity chromatography (Ni(II)-IMAC).

5.1.7 Columns

HisTrap™ HP 5 mL column	GE Healthcare, Munich, Germany
HiLoad™75 16/60 Superdex™75 *prepgrade*	GE Healthcare, Munich, Germany
HiLoad™75 26/60 Superdex™75 *prepgrade*	GE Healthcare, Munich, Germany
Superdex™ 75 10/300 GL	GE Healthcare, Munich, Germany
Superdex™ 200 10/300 GL	GE Healthcare, Munich, Germany

5.1.8 Kits

Molecular Biology Kits

E.Z.N.A.® Gel Extraction Kit	Omega bio-tek, Norcross, GA, U.S.A.
E.Z.N.A.® Cycle Pure Kit	Omega bio-tek, Norcross, GA, U.S.A.

E.Z.N.A.® Plasmid Mini Kit I	Omega bio-tek, Norcross, GA, U.S.A.
FavorPrep™ Gel/PCR Purification Kit	Favorgen, Vienna, Austria
FavorPrep™ Plasmid DNA Extraction Mini Kit	Favorgen, Vienna, Austria
QIAquick Gel Extraction Kit	Qiagen, Hilden, Germany
QIAquick PCR Purification Kit	Qiagen, Hilden, Germany
QIAprep Spin Miniprep Kit	Qiagen, Hilden, Germany
Crystallographic Kits	
JCSG+, JCSG Core I, II, III and IV	Qiagen, Hilden, Germany
MIDAS™	Molecular Dimensions, Suffolk, U.K.

5.2 Cloning and Plasmid Construction

5.2.1 Preparation of *P. aeruginosa* PAO1 genomic DNA

P. aeruginosa PAO1 were grown in 5 mL LB medium overnight at 37 °C. 1.5 mL of this culture was pelleted at 15700×g for 1 min, the supernatant was discarded and the cells were resuspended in Genomic DNA buffer (Section 5.1.1). The cell suspension was incubated at 37 °C for 1.5 hours. After addition of 250 µL chloroform and 250 µL phenol, the suspension was vortexed for 30 sec and pelleted at 15700×g for 1 min. The aqueous phase was then removed to another tube. These four steps were repeated four times. In the next step, 500 µL of chloroform were added. The suspension was vortexed and spun down at 15700×g for 1 min. The aqueous phase was removed to another tube. An appropriate volume of sodium acetate solution was added to a concentration of 0.3 M, as well as 1 mL of ethanol. The genomic DNA was then precipitated at -20 °C overnight. On the next day, the DNA was pelleted at 15700×g and 4 °C for 5 min. The DNA was washed with 70% (v/v) ethanol and resuspended in 100 µL of sterile, PCR grade ddH$_2$O.

5.2.2 Polymerase Chain Reaction

Polymerase chain reaction (PCR) is used to amplify target DNA. The ORFs encoding the proteins used in this study were amplified in PCR reactions from plasmid DNA or *P. aeruginosa* PAO1 genomic DNA. *Phusion* DNA Polymerase and *KAPAHiFi* Polymerase

(Peqlab) were used for amplification of DNA fragments for cloning, while *Taq* DNA Polymerase was used for colonyPCR (cPCR) reactions (Table 4).

Table 4: Pipetting scheme for PCR reactions. If using genomic DNA, more template DNA is required. DMSO is required because of the high GC content of *P. aeruginosa* genomic DNA.

	Phusion Polymerase	*KAPAHiFi* Polymerase	*Taq* Polymerase	final conc.
Template DNA	X µL	X µL	1 µL*	<200 ng/25 µL
Primer s	0.5 µL	0.75	0.5 µL	2-3 µM
Primer as	0.5 µL	0.75	0.5 µL	2-3 µM
dNTPs	0.5 µL	0.75	0.5 µL	200-300 µM
Buffer	5 µL GC Buffer	5 µL GC Buffer	2.5 µL Thermo Pol	1x
ddH$_2$O	to 25 µL	to 25 µL	17.25 µL	
DMSO	1.5 µL	0.5-1.5 µL	-	≤ 6%
Polymerase	0.125 µL	0.5 µL	0.25 µL	

*1 µl of a suspension of a single bacterial colony in 5 µL of ddH$_2$O.

The DNA polymerase was added last to the reaction mixture. Subsequently, the PCR programs (Table 5) were started. Denaturation, annealing and elongation were performed as cycles of 25 to 35 repetitions.

Table 5: Program scheme for PCR reactions. 25-35 cycles of denaturation, annealing and elongation were performed during PCR. T_m= melting temperature of the oligonucleotides; kbp = kilo base pairs.

		Phusion Polymerase		*KAPAHiFi* Polymerase		*Taq* Polymerase	
		T / °C	*t / s*	*T / °C*	*t / s*	*T / °C*	*t / s*
	Initial Denaturation	98	180	95	180	95	180
25 - 35	Denaturation	98	15	98	20	95	20
	Annealing	T_m- 5	20	T_m- 5	15	T_m- 5	20
	Elongation	72	30 / kbp	72	30 /kbp	68	60 /kbp
	Final Elongation	72	300	72	300	68	300
	Storage	8	∞	8	∞	8	∞

After PCR, the samples were mixed with DNA loading buffer and analyzed using agarose gel electrophoresis according to 5.2.4.

5.2.3 Gradient Polymerase Chain Reaction (Gradient PCR)

If standard PCR reaction was not successful in amplifying an ORF from *P. aeruginosa* genomic DNA, gradient PCR was used to amplify the target ORF. *Phusion* DNA Polymerase (NEB) and *KAPAHiFi* Polymerase (Peqlab) were used for amplification of these ORFs. Pipetting schemes are described in Table 6.

Table 6: Pipetting scheme for PCR reactions. If using genomic DNA, more template DNA is required. DMSO is required because of the high GC content of *P. aeruginosa* genomic DNA.

	Phusion Polymerase	*KAPAHiFi* Polymerase	final conc.
Template DNA	X µL	X µL	<200 ng/25 µL
Primer s	0.5 µL	0.75	2-3 µM
Primer as	0.5 µL	0.75	2-3 µM
dNTPs	0.5 µL	0.75	200-300 µM
GC Buffer	5 µL	5 µL	1x
ddH_2O	to 25 µL	to 25 µL	
DMSO	1.5 µL	0.5-1.5 µL	≤ 6%
Polymerase	0.125 µL	0.5 µL	

The PCR programs (Table 7) were started immediately after addition of the polymerase into the reaction mixture. Denaturation, annealing and elongation were performed as cycles of 10 (gradient step) and 20 (standard PCR step) repetitions.

Table 7: Program scheme for gradient PCR reactions. 10 cycles of gradient steps and 20 cycles of standard PCR steps were performed.

		Phusion Polymerase		*KAPAHiFi* Polymerase	
		T / °C	*t / s*	*T / °C*	*t / s*
	Initial Denaturation	98	180	95	180
10	Denaturation	98	15	98	20
	Annealing	T_m-5 to T_m+5	20	T_m-5 to T_m+5	15
	Elongation	72	30 / kbp	72	30 /kbp
20	Denaturation	98	15	98	20
	Annealing	T_m- 5	20	T_m- 5	15
	Elongation	72	30 / kbp	72	30 /kbp
	Final Elongation	72	300	72	300
	Storage	8	∞	8	∞

After PCR, the samples were mixed with DNA loading buffer and analyzed using agarose gel electrophoresis according to 5.2.4.

5.2.4 Agarose Gel Electrophoresis

Nucleic acids samples from PCR were analyzed using agarose gel electrophoresis. In this analysis, 1% agarose gels (1 g of agarose per 100 mL 1x TAE buffer) were supplemented with 2 µL of ethidium bromide solution. Samples were mixed with 6x DNA loading buffer and loaded on the gel. Voltage of 35 V or 70 V, depending on the size of the agarose gel, was applied. The DNA fragments migrate in the agarose gel and are separated according to their size. After electrophoresis, gels were inspected under UV light. DNA fragments were compared to a molecular weight standard, cut out with a scalpel and purified using a DNA purification kit (see Section 5.1.8) according to the manual of the manufacturer.

5.2.5 DNA Restriction

PCR products and plasmids were treated with endonuclease restriction enzymes (New England Biolabs) to generate 3'-overhangs for ligation reactions. Plasmids were additionally treated with Antarctic Phosphatase (New England Biolabs) to dephosphorylate the 5'-ends of the DNA and decrease the probability for religation. Table 8 depicts the pipetting schemes for digestion reactions of PCR products and plasmid DNA. Buffers were chosen according to the manual of the manufacturer.

Table 8: Pipetting schemes for endonuclease restriction digest of linear and circular dsDNA.

	PCR products	Plasmid DNA
Template DNA	X µL (700 ng)	X µL (1-2 µg)
10x BSA	5 µL	5 µL
10x NEB Buffer	5 µL	5 µL
Restriction Enzyme I	1 µL	1 µL
Enzyme II	1 µL	1 µL
Antarctic Phosphatase (AP)	-	1 µL
10x AP Buffer	-	5 µL
ddH$_2$O	to 50 µL	to 50 µL

The DNA was digested for 30 min (PCR products) or 1 hour (plasmid DNA) at 37 °C. Finally, the enzymes were denatured at 65°C for 20 min (80 °C for NcoI) and purified according to 5.2.6.

5.2.6 Purification of Linear DNA

Purification of linear DNA, e. g. from enzymatic digest, was performed using a PCR Purification Kit (see Section 5.1.8) according to the protocols supplied. DNA was either eluted with ddH_2O or the supplied elution buffer.

5.2.7 Quantification of Nucleic Acids

Nucleic acid concentrations were determined using a NanoDrop 1000 or NanoDrop 2000 (Thermo Scientific). Solutions were measured directly at a UV absorption wavelength of λ = 260 nm. A UV absorption of 1 was assumed to be equal to 50 ng/µL of double stranded DNA. Concentrations were calculated accordingly.

5.2.8 Ligation

After restriction of PCR products and vectors by endonucleases, PCR products (inserts) and plasmids were ligated using T4 DNA Ligase in order to generate the desired expression vectors. The reactions (Table 9) were setup using an insert:plasmid ratio between 3:1 and 5:1.

Table 9: Pipetting scheme for ligation reactions.

	Ligase Reaction I	Ligase Reaction II
Insert	6 µL	x µL (3-5 equivalents)
Plasmid	2 µL	x µL (1 equivalent)
Ligase Buffer (NEB)	1 µL	1 µL
T4 DNA Ligase	1 µL	1 µL
ddH_2O	to 10 µL	to 10 µL

Ligase reaction proceeded at 16 °C for either for one, two, four or 16 hours. After ligation, T4 DNA Ligase was inactivated at 65 °C for 20 min. The reaction mixture was used for transformation into *E. coli* cells without further purification (Section 5.2.13).

5.2.9 Preparation of Circular DNA

Plasmid DNA was purified from *E.coli* strains XL1-Blue, TOP10 or DH5α using a Plasmid Purification Kit (see Section 5.1.8). Purification was performed according to the protocols supplied. DNA was eluted with ddH$_2$O or the supplied elution buffer.

5.2.10 DNA Sequencing

In order to verify the correct sequence of the constructed plasmids, DNA sequencing analysis was performed at MWG Biotech/Eurofins Genomics (Ebersberg, Germany) or by the Genome Analytics Facility (GMAK) of the Helmholtz Centre of Infection Research (Braunschweig, Germany) according to the respective protocols. Sequencing results were analyzed in pairwise alignments using *ClustalW2* (European Molecular Biology Laboratory, EMBL) (Larkin *et al.*, 2007).

5.2.11 Preparation of *E.coli* Chemically Competent Cells

Chemically competent *E. coli* strains were produced for molecular biology and protein production techniques. Magnesium competent and rubidium competent cells were prepared by Norbert Grillenbeck at the University of Bayreuth and Claudia Hanko at the Helmholtz Centre for Infection Research, respectively. *E. coli* cells were prepared according to standard protocols, using antibiotics only if the cells were carrying an additional plasmid (e.g. chloramphenicol with *E. coli* BL21(DE3)Codon+ cells). Cells were flash frozen in liquid nitrogen as 100 µL aliquots and stored at -80 °C until further usage.

5.2.12 Preparation of *P. aeruginosa* Magnesium-Competent Cells

Magnesium-competent *P. aeruginosa* PAO1 were prepared for protein production in spot plate assays. *P. aeruginosa* PAO1 were cultivated on LB agar at 37 °C overnight. A

20 mL LB culture was inoculated with a single colony and cultivated at 37 °C under vigorous shaking to late stationary phase. Cells were pelleted at 15700×g and LB medium was removed. The cells were then suspended in 5 mL of ice cold 0.1 M $MgCl_2$ for 30 min at 4 °C. After centrifugation and removal of $MgCl_2$, cells were suspended in 2.5 mL of 0.15 M $MgCl_2$ and 15% (*v/v*) glycerol. Aliquots of 100 µL were flash frozen in liquid nitrogen and stored at -80 °C.

5.2.13 Heat Shock Transformation of *E. coli* and *P. aeruginosa* PAO1

For plasmid amplification or protein production, 50 µL of *E. coli* chemically competent cell suspension were mixed with 50 ng of plasmid DNA and incubated on ice for 30 min. A 90 s heat shock was performed at 42 °C and 1 mL of LB medium was added to the cells. After shaking for 10 min (if resistant to ampicillin) or 30 min (if resistant to chloramphenicol), cells were pelleted at 15700×g for 1 min and plated on LB agar containing appropriate antibiotics.

P. aeruginosa PAO1 chemically competent cells were treated in the same manner as *E. coli*. They were shaken for 30 min (gentamicin) in LB medium at 37 °C before they were plated on LB agar containing appropriate antibiotics.

5.2.14 Protein Production

All proteins used in this study were produced in *E. coli*. An overnight culture of LB medium supplemented with antibiotics was inoculated with a single colony from an agar plate with freshly transformed cells. Cells were cultivated at 37 °C and under vigorous shaking to an optical density at 600 nm (OD_{600}) of 0.6. The temperature was then set to 20 °C and cells were cultivated to an OD_{600} of 0.8, at which gene expression was induced with 0.5 mM isopropyl-thio-β-D-galactopyranosid (IPTG). Cells were cultivated at 20 °C overnight under vigorous shaking (120 rpm) and harvested by centrifugation at 5000×g for 10 min at 4 °C. If not used immediately, the cells were flash frozen in liquid nitrogen and stored at -80 °C. Variances in the protein production protocol are listed in the table below. All proteins produced during this study and are listed in Table 10. PA3127_1-114 and PA3127_123-267 were produced by Monika Popp (Popp, 2013).

Table 10: Proteins produced in this study. PA3127_1-114 and PA3127_123-267 were produced by M. Popp (Popp, 2013). The *E. coli* strain and plasmid used for production are indicated.

Protein	*E. coli* strain	Plasmid	Citation
His_6-PA0803	Rosetta 2 pLysS	pET19m_PA0803	Yu, 2009, 16h at 20 °C ON
His_6-PA0880	BL21 (DE3)	pET19m_PA0880	this study
His_6-PA1353	BL21 (DE3) Codon+	pET19m_PA1353	Kalawy-Fansa, 2010
His_6-PA1354	BL21 (DE3)	pET19m_PA1354	this study
His_6-PA1672	BL21 (DE3) pLysS	pET19m_PA1672	this study, Weber, 2011
His_6-PA3127	Rosetta 2 pLysS	pET19m_PA3127	this study, Bittner, 2012
His_6-PA3127_1-114	BL21 (DE3) Codon+	pET19m_PA3127_1-114	Popp, 2013
His_6-PA3127_123-267	BL21 (DE3) Codon+	pET19m_PA3127_123-267	Popp, 2013
His_6-PA4183	BL21 (DE3) pLysS	pET19m_PA4183opt	this study
His_6-PA4183_16-155	BL21 (DE3) Codon+	pET19m_PA4183opt_16-155	this study
His_6-PA4183_32-155	BL21 (DE3)	pET19m_PA4183opt_32-155	this study
His_6-PA4518	Rosetta 2 pLysS	pET19m_PA4518	this study
His_6-PA4641	BL21 (DE3) Codon+	pET19m_PA4641	Kalawy-Fansa, 2010

5.3 Protein Purification and Characterization

5.3.1 Protein Purification

Cell Disruption

Cells were thawed if previously frozen and then resuspended in an appropriate volume of buffer A. If necessary, the suspension was supplied with DNAse K and magnesium chloride. The suspension was homogenized by stirring and filtrated. Cells were lysed by sonication or in a microfluidizer according to the instructions of the manufacturer. To separate the soluble proteins from the insoluble fraction and cell debris, a centrifugation step was performed at 20000×g and 4 °C for 60 min. The supernatant of the centrifugation containing the soluble target protein was pooled for further purification. Samples for sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were taken from the cell lysate and supernatant.

Ni(II) Immobilized Metal Ion Affinity Chromatography (Ni(II)-IMAC)

A *HisTrap™ HP* column loaded with Ni(II) and primed with buffer A was connected to an *ÄKTAprime* FPLC system which previously had been flushed with buffers A and B. The whole system was primed with buffer A until a stable baseline at an absorption wavelength of $\lambda = 280$ nm (A_{280}) was reached. The supernatant from the previous centrifugation step was applied to the *HisTrap™ HP* column at a flow rate of 1-3 mL/min and a maximum pressure of 0.3 MPa. The column was washed with 2% of buffer B to elute unspecifically bound proteins. After that, a 120 mL gradient of 2-50% buffer B (10-250 mM imidazole) was applied to the column in order to elute the pure target protein. The elution was collected in 2 mL fractions and samples were analyzed by SDS-PAGE.

Dialysis and TEV Cleavage

The fractions pooled from Ni(II)-IMAC purification containing the soluble target protein were filled into a 10 kDa molecular weight cut-off (*MWCO*) dialysis bag. One equivalent of TEV protease was added to 20 equivalents of target protein (calculated as molar ratio). The dialysis bag was sealed and was transferred into 4.5 L of SEC buffer supplemented with 1 mM βME. Dialysis was performed under continuous stirring at 4 °C overnight.

Reverse Ni(II) Immobilized Metal Ion Affinity Chromatography (revNi(II)-IMAC)

The protein solution from dialysis and TEV cleavage was used in a revNi(II)-IMAC which was performed similarly to Ni(II)-IMAC. However, SEC buffer (instead of buffer A) was used to prime the *HisTrap™ HP* column and the *ÄKTAprime* FPLC system. Since the His_6-tag of the target protein had been cleaved during dialysis, the target protein was present in the flow-through of the *HisTrap™ HP* column. Fractions of 2 mL of the flow-through were collected and analyzed by SDS-PAGE. His_6-tagged TEV protease as well as other impurities such as non-cleaved target protein bound to the *HisTrap™ HP* column and were therefore eluted in a 100 mL gradient of 0-50% buffer B (0-250 mM imidazole). The elution fractions were collected and analyzed by SDS-PAGE. This step is very important due to the three species of βαβββ-module proteins present in solution: completely cleaved dimer, dimer consisting of one cleaved and one uncleaved protein and completely uncleaved dimer.

Preparative Size Exclusion Chromatography

A homogenous protein solution of high purity is crucial for protein crystallization. For this purpose, size exclusion chromatography (SEC) was always performed as the last protein purification step. Proteins were loaded onto a HiLoad™75 16/60 Superdex™75 *prepgrade* or HiLoad™75 26/60 Superdex™75 *prepgrade* column equilibrated in SEC buffer and connected to an *ÄKTAprime+* or *ÄKTApurifier* High-Performance Liquid Chromatography (HPLC) system (GE Healthcare). Size exclusion chromatography was performed at a maximum column pressure of 0.3 MPa or the maximum flow rate possible according to the manufacturer's manual. Proteins were eluted isocratically in SEC buffer and collected in 2 mL fractions. Samples were analyzed by SDS-PAGE. Fractions containing pure homogenous protein were pooled and concentrated by ultrafiltration. All size exclusion chromatography experiments were performed at 4-10 °C. The molecular weight of the target protein was calculated from a standard calibration curve obtained from the retention times of standard proteins with known molecular weight (Bio-Rad).

Variances in the Purification Protocol

All proteins used in this study (Table 11) were purified with an N-terminal His$_6$-tag that was not cleaved in every case.

Table 11: Proteins purified and/or used in this study. PA3127_1-114 (NTD) and PA3127_123-267 (CTD) were received from M. Popp (Popp, 2013).

Protein	His$_6$-tag cleaved?	Crystallization	Citation
His$_6$-PA0803	no	yes	Yu, 2009
PA0880	yes	yes	this study
His$_6$-PA1353	no	-	Kalawy-Fansa, 2010
His$_6$-PA1354	no	yes	this study
PA1672	yes	yes	this study, Weber, 2011
PA3127	yes	yes	this study, Bittner, 2012
PA3127_1-114	yes	-	Popp, 2013
His$_6$-PA3127_123-267	no	yes	Popp, 2013
PA4183	insoluble	-	this study
PA4183_16-155	yes	yes	this study
PA4183_32-155	yes	yes	this study
PA4518	yes	yes	this study
His$_6$-PA4641	no	-	Kalawy-Fansa, 2010

If the His_6-tag was cleaved, all purification steps described above were performed. In cases where His_6-tag cleavage was not necessary, neither dialysis nor reverse Ni(II)-IMAC were performed. After purification on Ni(II)-IMAC, the protein solutions were concentrated as described in 5.3.3 and applied to a size exclusion chromatography column. All proteins, except PA3127_1-114 and PA3127_123-267, were purified during this study.

5.3.2 SDS-PAGE

Sodium-Dodecyl-Sulfate Gel Electrophoresis (SDS-PAGE) was used to verify the presence and analyze the purity of proteins during purification steps. Discontinuous gels consisting of a stacking gel (5% polyacrylamide) and a resolving gel (16% or 18% polyacrylamide) were prepared following standard protocols (Laemmli, 1970). Samples were prepared by adding 5x SDS-PAGE sample buffer to the protein solution and denaturing the sample at 95 °C for 5 min. After centrifugation at 15700×g, samples were run on the gel at a voltage of 200 V in running buffer until the dye front had reached the end of the gel. After SDS-PAGE, the gel was washed with water, stained in staining solution containing Coomassie Brilliant Blue and destained in destaining solution or in H_2O.

5.3.3 Ultrafiltration

During and after purification, proteins were concentrated to a desired volume and concentration using ultrafiltration. A *Millipore Amicon Ultra MWCO* or *PALL Macrosep® Advance Centrifugal Device* of 10 kDa MWCO was equilibrated with the appropriate buffer by centrifugation at 2900×g for 5 minutes. The buffer was removed and the protein solution was applied to the membrane. The protein solution was concentrated to a desired concentration by centrifugation at 2900×g.

5.3.4 Protein Concentration Determination

Protein concentrations were determined directly by UV absorption at a wavelength of $\lambda = 280$ nm using a *NanoDrop 1000* or *NanoDrop 2000* (Thermo Scientific). The theoretical molecular weight and extinction coefficients of the proteins (Table 12) were calculated according to the appropriate protein sequence using EXPASY ProtParam (Gasteiger *et al.*, 2003). Protein concentrations were then determined using Lambert-Beer's law

$$A = c \cdot d \cdot \varepsilon_{280} \qquad (22)$$

when A is the absorption, c is the concentration (in M), d is the path length (in cm) and ε_{280} is the extinction coefficient at λ = 280 nm (in $M^{-1}cm^{-1}$).

Table 12: Molecular weight (in kDa) and extinction coefficients ε_{280} used for concentration determination of the proteins used in this study.

Protein	Molecular weight / kDa	Extinction coefficient ε_{280}
His_6-PA0803	17.94	27055
His_6-PA0880	15.75	4470
PA0880	13.87	2980
His_6-PA1353	15.28	22710
His_6-PA1354	18.03	13980
His_6-PA1672	15.49	5960
PA1672	13.61	4470
His_6-PA3127	31.47	45045
PA3127	29.60	43555
PA3127_1-114	12.74	11585
His_6-PA3127_123-267	18.20	33460
His_6-PA4183_16-155	17.46	11460
PA4183_16-155	15.58	9970
His_6-PA4183_32-155	15.68	5960
PA4183_32-155	13.80	4470
His_6-PA4518	17.22	32555
PA4518	15.35	31065
His_6-PA4641	17.28	24325

5.3.5 Analytical Size Exclusion Chromatography

To determine the oligomerization state of a target protein, analytical size exclusion chromatography was performed. Analytical size exclusion chromatography proceeded similarly to preparative size exclusion chromatography described in section 5.3.1. Instead of a preparative column, Superdex™ 75 10/300 GL or Superdex™ 200 10/300 GL columns (24 mL column volume) connected to an *ÄKTA* FPLC system was equilibrated with SEC buffer. Furthermore, only 1 mg of protein was loaded onto the column. The elution profile was compared to the one of a molecular weight standard (Bio-Rad), which was used to calculate the molecular weight and oligomeric state of the protein of interest.

5.3.6 Bleomycin N-Acetyltransferase Activity Test using HPLC-MS

Bleomycin was acquired as Bleomedac® from Medac GmbH (Wedel, Germany) as a mixture of bleomycin A_2 and B_2. HPLC and Mass spectrometry (MS) analysis were performed to analyze the enzymatic reactions catalyzed by PA3127 and PA3128. The assay samples were prepared using each 260 µM acetyl-Coenzyme A (AcCoA), 200 µM NADPH, 200 µM $NADP^+$, 100 µM bleomycin, 5 µM PA3128 and 5 µM PA3127 in a volume of 500 µL of 50 mM sodium acetate pH 3.6, 4.6 or 5.6 or in 50 mM BisTris pH 5.8, 6.5 or 7.2. The reactions proceeded at 37 °C for 5 min, 15 min, 30 min, 1 hour or 2 hours and stopped by heating the sample to 80 °C. The precipitated proteins were pelleted by centrifugation at 15700×g. The supernatant was freeze-dried in liquid nitrogen and lyophilized overnight. The samples were measured according to protocols developed by Ulrike Beutling in the laboratory of Prof. Dr. Mark Brönstrup, Department of Chemical Biology at the Helmholtz Centre for Infection Research in Braunschweig, Germany.

5.4 Protein Crystallization and Structure Determination

5.4.1 Determination of Crystallization Conditions

In order to identify an initial crystallization hit, the vapor diffusion method was used with a sitting drop set up. Crystallization trials were set up manually or using liquid handling systems (*Phoenix*, Art Robbins Instruments; *Honeybee 961*, Zinsser Analytics*; OryxNano*, Douglas Instruments). When using *Honeybee 961* or *OryxNano*, screens were set up on Intelli-Plate™ 96-3 LVR plates, while *SWISSCI MRC 2-Well* or *3-Well* crystallization plates were used with the *Phoenix*. When using *OryxNano*, drops were composed of 0.3 µL of reservoir and 0.3 µL of protein solution, when using *Honeybee 961* or *Phoenix*, 0.2 µL of reservoir and 0.2 µL of protein solution were pipetted. When setting up plates manually, crystallization trials were prepared as 0.8 µL + 0.8 µL drops on *SWISSCI MRC 2-Well* or *3-Well* crystallization plates. In all cases, the amount of reservoir solution ranged from 35 to 70 µL (Table 13). All proteins were screened using the JCSG Core suites I, II, III and IV (Qiagen). PA3127_123-267 was additionally screened with the Midas™ screen (Molecular Dimensions).

Table 13: Proteins used in crystallization trials. Concentration of the protein, crystallization robot, as well as the reservoir volume are listed. PYO = pyocyanin, DTT = dithiothreitol, AcCoA = acetyl-coenzyme A, CoA = coenzyme A, Bm = bleomycin, I3C = 5-Amino-2,4,6-triiodoisophthalic acid.

Protein	Concentrations / mg/mL	Robot	Volume of reservoir / µL
His$_6$-PA0803	20 (± 1 mM PYO)	*Phoenix*	40
His$_6$-PA0880	20 and 15 (+ 2 mM DTT)	*Phoenix*	40
PA0880	24, 20 and 15 (+ 2 mM DTT)	*Phoenix*, *Honeybee*	40, 70
His$_6$-PA1354	30 and 22.5	*Phoenix*	40
PA1672	20 and 15	set up manually	40
PA3127	20, 15, 10, 8, 4 (± 2 mM AcCoA / CoA / Bm / CoA + Bm / I3C)	*Phoenix, Honeybee*	40, 70
His$_6$-PA3127_123-267	12.6 (± 2 mM AcCoA)	*Honeybee*	70
PA4183_16-155	25 and 15	*Phoenix*	35
PA4183_32-155	30 and 15	*OryxNano* (at 4 °C)	70
PA4518	20 and 15 (+ 2 mM DTT)	set up manually	40

All crystallization trials were set up at 20 °C, sealed and stored at 20 °C in an automated imaging system (Formulatrix). In this system, crystallization screens were imaged regularly according to predefined schedules.

5.4.2 Crystal Optimization

Initial crystallization hits were reproduced and optimized in a vapor diffusion hanging drop setup using *Linbro* 24-well plates with 22 mm glass slides or *VDX*m 24-well plates and 18 mm glass slides. In these optimization plates, parameters such as the protein concentration, the drop size, the concentration of the precipitant or the pH of the reservoir were varied to obtain better diffracting crystals. In some cases, optimization plates are also used for soaking experiments with ligands or heavy atom-containing solutions. Usually, 500 µL of mother liquor were used to equilibrate the crystallization drops usually composed of 1 µL of reservoir and 1 µL of protein solution. Plates were stored at 20 °C and inspected under a microscope for crystals once per day.

Table 14: Initial hits and final crystallization conditions of the proteins investigated in this study. MPD = 2-Methyl-2,4-pentanediol, MES = 2-(N-morpholino)ethanesulfonic acid.

Protein	(Best) Initial Hit	Final Condition
His_6-PA0803	JCSG Core IV H6	0.8 M sodium dihydrogen phosphate, 1.2 M dipotassium hydrogen phosphate, 0.1 M sodium acetate pH 4.5, 2.5 mM PYO, 2.5% (*v/v*) DMSO
PA0880	JCSG Core II F3	0.1 M sodium acetate pH 4.5, 50% (*v/v*) ethylene glycol, 5% (*w/v*) PEG 1000
His_6-PA1354	JCSG Core I G2	20% (*w/v*) MPD, 0.1 M sodium acetate pH 5
PA1672	JCSG Core I H2	0.1 M sodium acetate pH 4.5, 20% (*w/v*) PEG 3350
PA3127	JCSG Core III F9	0.1 M MES pH 5.6, 0.8 M ammonium sulfate
PA3127_123-267	Midas A1	0.1 M HEPES/NaOH pH 6, 50% (*v/v*) polypropylene glycol 400, 5% (*v/v*) DMSO
PA4183_16-155	JCSG Core I B1	0.1 M HEPES/NaOH pH 7.2, 0.2 M sodium chloride, 18% (w/v) PEG 3350
PA4518	JCSG Core II H11	0.1 M sodium citrate pH 3.6, 10% (*v/v*) MPD

5.4.3 Cryoprotection

Crystallographic data usually is collected at cryogenic temperatures to reduce radiation damage introduced into the protein crystal by X-rays. In order to avoid the formation of amorphous ice at these temperatures (usually 100 K), protein crystals were soaked with a cryoprotecting solution composed of the mother liquor and high precipitant concentrations, for instance glycerol, sugars, oil or small molecular weight polyethylene glycols (PEGs) (Table 15).

If ligands were present in the crystallization solution, they were added to the cryoprotecting solution at the same concentration, for instance 2.5 mM pyocyanin was added in the case of PA0803. After cryoprotection, crystals were isolated in nylon loops of the appropriate size and flash cooled in liquid nitrogen for diffraction experiments.

Table 15: Cryoprotectants used for the diffraction experiments on protein crystals. The cryoprotecting solution always consists of the mother liquor, a cryoprotecting agent and ligands.

Protein	Cryoprotectant
His_6-PA0803	15% (*w/v*) glycerol
PA0880	-
His_6-PA1354	-
PA1672	5% (*w/v*) PEG 3350, 15% (v/*v*) glycerol
PA3127	25% (*v/v*) glycerol
His_6-PA3127_123-267	-
PA4183_16-155	12% (w/v) PEG 3350
PA4518	15% (*v/v*) MPD

5.4.4 Diffraction Experiments and Data Collection

After crystals were obtained, isolated and cryoprotected, they were tested for diffraction on X-ray home sources. Test images were collected using monochromatic X-rays at the Cu-K_α edge (λ = 1.54 Å) together with either an *IµS* sealed tube anode (Incoatec) and a *mar345* (Rayonix) image plate detector, or a rotating anode with an *R-AXIS IV++* (Rigaku) image plate detector or a *Saturn 944+* (Rigaku) CCD detector. Well-diffracting crystals were sent to a synchrotron for data collection. All datasets were collected on beamline 14.1 of the BESSY II Synchrotron (Berlin Electron Storage Ring Society for Synchrotron Radiation, Helmholtz Centre Berlin (HZB), Berlin, Germany) at cryogenic temperatures and at a wavelength of λ = 0.9184 Å (Table 16).

Table 16: Data collection strategies for the crystals used for data collection in this study.

Protein	Beamline	Detector	Wavelength λ / Å	Images	Oscillation / °
His_6-PA0803 + PYO	BESSY II 14.1	MarCCD Chess	0.9184	200	0.5
PA0880	BESSY II 14.1	MarCCD Chess	0.9184	100	1
His_6-PA1354	BESSY II 14.1	MarCCD Chess	0.9184	105	1
PA1672	BESSY II 14.1	MarCCD Chess	0.9184	100	1
PA3127 relaxed cell	BESSY II 14.1	MarCCD Chess	0.9184	110	1
PA3127 dehydrated cell	BESSY II 14.1	PILATUS 6M	0.9184	1900	0.1
His_6-PA3127_123-267	BESSY II 14.1	PILATUS 6M	0.9184	3600	0.1
PA4518	BESSY II 14.1	MarCCD Chess	0.9184	100	1

5.4.5 Data Processing

All data were indexed with *XDS* (Kabsch, 2010) and integrated and scaled using *Aimless* (Evans, 2006, 2011). For every protein and space group, a reference dataset was created. This reference dataset was used to assign the identical free R flags to all datasets of this protein and space group. By doing so, the statistical parameters of all structures can be compared more easily. The final structure factor file was used for refinement and improvement of the corresponding protein model.

5.4.6 Phase Determination and Structure Determination

The crystal structures of the proteins investigated in this part of the study were determined via MR. This was possible because structures of similar fold and sequence are available from the PDB. Structure solution proceeded either in *PHASER* (McCoy *et al.*, 2007) from the *CCP4* program suite (Winn *et al.*, 2011), or in *PHASER* from the *Phenix* program suite (Adams *et al.*, 2011). If these approaches were unsuccessful, MR was tried using the MR engine *BALBES* (Long *et al.*, 2007). An overview on the search models used in MR and their sequence identities to the target proteins is given in Table 17.

Table 17: Structures determined in this study. All structures were solved by MR with homologues proteins, either using *PHASER* from *CCP4* or *Phenix*, or using the MR engine *BALBES*.

Protein	Structure Solution	Model	sequence identity / %
His$_6$-PA0803	rigid body refinement	PA0803 (Yu, 2009)	100
PA0880	*PHASER*, CCP4	3HNQ, dimer	66
PA1672	*PHASER*, CCP4	2R6U, chains BD, polyALA	28
PA3127	*PHASER*, Phenix	PA3127_1-114, dimer	43
		PA3127_123-267, dimer	54
His$_6$-PA3127_123-267	BALBES	3JVN, monomer	11
PA4518	*PHASER*, CCP4	3BQX, dimer	37

5.4.7 Model Building and Refinement

If a structure solution is obtained by MR or phasing, the structural model has to be rebuilt and improved by iterative building and refinement cycles. These include the placement of water molecules and ions as well as ligands into the protein structure. For this purpose, model building was done in *COOT* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010) and refinement was carried out with *REFMAC5* (Murshudov *et al.*, 2011) and with *phenix.refine* (Afonine *et al.*, 2012). If ligands were placed in the protein structure, coordinate files and chemical restraints files were downloaded from the PDB and then prepared for refinement using *eLBOW* (Moriarty *et al.*, 2009) from the *Phenix* program suite (Adams *et al.*, 2011). In case ligands were not available in the PDB, coordinate files and chemical restraints files were prepared with *ProDRG* (Schüttelkopf & van Aalten, 2004) and *eLBOW* (Moriarty *et al.*, 2009).

5.4.8 Structural Analysis and Validation

All structures were validated using *MolProbity* (Chen *et al.*, 2009) and the Validation Server of the PDB (Berman *et al.*, 2000). Figures were prepared with *PyMOL* (DeLano, 2002), *Chimera* (Pettersen *et al.*, 2004) and *ChemBioDraw*.

5.5 Small Angle X-ray Scattering

5.5.1 Principles of Small Angle X-ray Scattering

Small angle X-ray scattering (SAXS) is a technique that is used to visualize the overall 3-dimensional structure and structural changes of biological macromolecules. The method reveals the size and shape of the electron envelope of proteins and protein complexes ranging from 5 kDa to 100 MDa, and complements X-ray crystallography, nuclear magnetic resonance (NMR) and electron microscopy (EM) (Putnam *et al.*, 2013).

In a SAXS experiment, a monodisperse protein solution is probed with an X-ray beam (λ around 1.5 Å) and the scattered X-rays are detected at low angle (θ = 0.1 - 10°) to the incident beam. The amplitude |**s**| of scattering vector **s** is

$$|\vec{s}| = 4\pi sin\theta/\lambda \tag{23}$$

where 2θ is the scattering angle between incident and scattered beam. At values of **s** < $1.3R_g$, the scattered intensities I_s can also be described by the Guinier approximation

$$I_{\vec{s}} = I_0 e^{-\vec{s}^2 R_g^2/3} \tag{24}$$

where I_0 is the intensities at $\theta = 0$ and R_g is the radius of gyration. The R_g is defined as a measure for the overall size of the molecule, meaning the average of square center-of-mass distances in the molecule. Both the R_g and the molecular weight MM_P of the molecule of interest can then be determined using the Guinier plot ($\ln(I_s)$ plotted against $\mathbf{s}^2$) by the I_0 intercept and the slope of the plot, respectively. The MM_P can also be determined by comparison of the forward scattering intensities $I_{0,st}$ and MM_{st} of a reference protein solution, for instance BSA (66.4 kDa), with the foward scattering $I_{0,P}$ of the protein of interest, using

$$\boldsymbol{MM_P = I_{0,P}/c_P \frac{MM_{st}}{I_{0,st}/c_{st}}} \tag{25}$$

where c is the concentration. Both the R_g and the molecular weight MM_P of the molecule are used in data processing, together with the excluded volume V (Porod's volume) of the hydrated particle which is computed using Porod's equation

$$\boldsymbol{V = 2\pi I_0 \int_0^\infty \vec{s}^2 I_{exp,\vec{s}} d\vec{s}} \tag{26}$$

In data processing, the scattered intensities I_s are first Fourier-transformed into 1-dimensional scattering curves. Then, the curves can be interpreted using the distance distribution function $p(r)$ of the particle

$$\boldsymbol{p(r) = 1/2\pi^2 \int_0^\infty I_s \vec{s}\vec{r} \sin \vec{s}\vec{r}\, d\vec{s}} \tag{27}$$

where r is a distance in the particle. The particle distance distribution function represents all paired sets of distances between the electrons of the molecule of interest. The function $p(r)$ yields an *ab initio* shape of the molecule of interest. This *ab initio* shape is used together with the scattering curves in order to reconstruct the low resolution 3-dimensional shell. This process is usually aided by X-ray crystallography or NMR structures of the proteins of interest. For this purpose, the protein is represented as a bead model from which a theoretical scattering curve can be calculated. The model is altered until the theoretical scattering curve fits to the observed scattering curve. An important quality indicator for this fitting process is the X-value

$$\boldsymbol{X^2 = \frac{1}{N} \sum_{j=1}^{N} \left(I_{exp,\vec{s}_i} - \eta I_{calc,\vec{s}_i}/\sigma_{\vec{s}_i}\right)^2} \tag{28}$$

where N is the number of experimental points, I_{exp,s_i} and I_{calc,s_i} are the experimental and the theoretical intensities, η is a scaling factor and σ_{s_i} is the experimental error. A X-value of between 1 and 5 usually indicates a high quality SAXS model.

5.5.2 Small Angle X-ray Scattering – Experiments and Data Evaluation

SAXS measurements were carried out on beamline BM29 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using a PILATUS 1M detector. Samples were prepared as a 100 µL dilution series of purified PA3127 at concentrations of 15, 5, 2.5, 1.25, 0.75 and 0.375 mg/mL in SEC buffer, which was used as a blank. Solution scattering patterns were collected at a sample – detector distance of 2.867 m and a wavelength of λ = 0.9919 Å, covering the range of momentum transfer $0.01 < \mathbf{s} < 0.45$ nm^{-1} (equation 23). In case of PA3127, the best sample quality and monodispersity (91.8%) was achieved at a protein concentration of 1.25 mg/mL. This sample was chosen for further analysis of the scattering data. First, the crystal structure of PA3127 (determined in this study) was submitted to the *PISA* server (Krissinel & Henrick, 2007) in order to determine the correct oligomeric assembly for electron shell calculation. All stable protein assemblies predicted by *PISA* were used for SAXS curve fitting in the program *CRYSOL* (Svergun *et al.*, 1995). The quality of each curve fit is indicated by the X-value (equation 27), when $\mathrm{X} = 1$ indicates the optimum fit of the structure and the electron shell computed from the recorded scattering data. Then, the program *PRIMUS* (Konarev *et al.*, 2003) was used to normalize and average the scattering data and subtract the buffer blank. Furthermore, *PRIMUS* evaluates the integral experiment parameters, for instance the radius of gyration $\mathrm{R_g}$, zero intensity $\mathrm{I_0}$ and molecular weight $\mathrm{MM_P}$ of the protein of interest and the excluded volume V of the hydrated particle. The same parameters were also computed using the program *GNOM* (Svergun, 1992) from the *ATSAS* package (Petoukhov *et al.*, 2012). *GNOM* also provides the maximum diameter $\mathrm{D_{max}}$ of the particle and evaluates the particle distance distribution function $\rho(\mathrm{r})$. Then, *GNOM* analyzes the one-dimensional scattering curves fitted by *CRYSOL.* These data were used for twenty runs of *ab initio shape* determination of the SAXS envelope by the program *DAMMIF* (Franke & Svergun, 2009). For this purpose, the oligomeric assembly is represented as a densely packed bead model, and *DAMMIF* uses simulated annealing to determine twenty *ab initio* models of the PA3127 SAXS envelope. These models were aligned and averaged using *DAMAVER* (Volkov & Svergun, 2003). In *DAMAVER*, first *damsel* compared all models composed in *DAMMIF* and identified the most probable one as well as outliers. Second, *damsup* aligned the models with the most probable one. Third, *damaver* averaged the models and computed a model probability map. In the last step, *damfilt* filters the averaged model at a certain cut-off volume and yields a final SAXS envelope. The SAXS envelope was superimposed with the correct PA3127 assembly using *SUPCOMB20* (Kozin & Svergun, 2001). The assembly and the final electron shell were visualized using *SITUS* (Wriggers, 2012) and *Chimera* (Pettersen *et al.*, 2004).

5.6 Isothermal Titration Calorimetry

5.6.1 Principles of Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a highly sensitive method to measure the interaction between two molecules. These molecules can either be two proteins or proteins and their small ligands. The actual signal detected in an ITC experiment is the power difference between the sample cell and the reference cell generated by molecules that absorb or generate heat upon interaction.

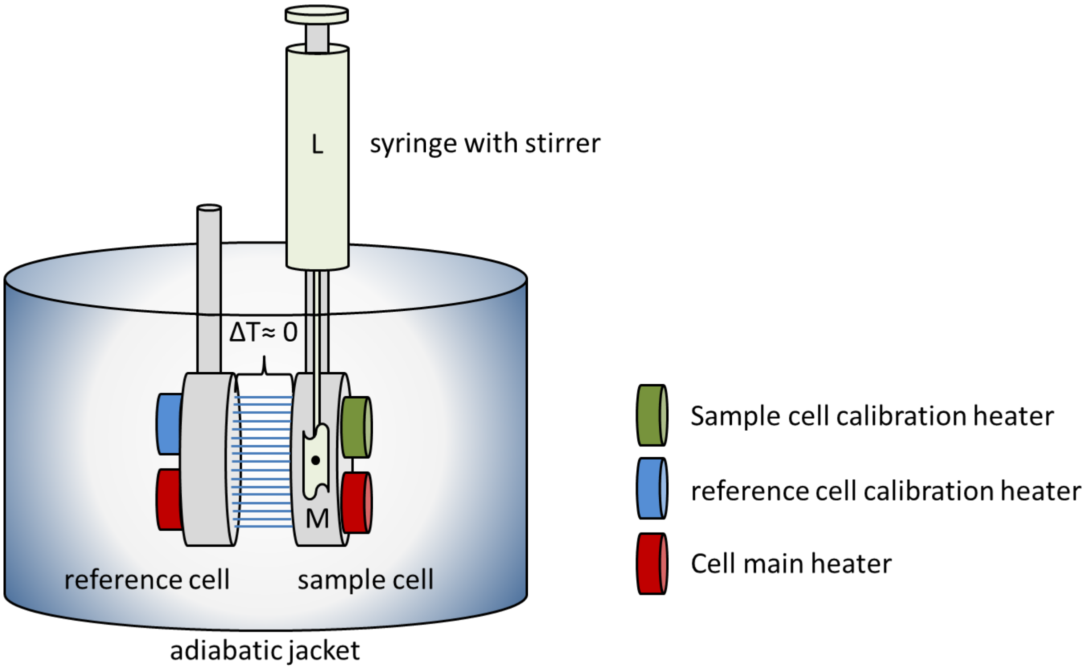


Figure 21: Experimental setup of isothermal titration calorimetry (ITC). The reference cell and the sample cell are enclosed in an adiabatic jacket and are constantly heated with the cell main heaters. The temperature difference ΔT between both cells is very small, but measurable. The protein solution (M) is loaded into the sample cell, and the ligand solution (L) is filled into the stirring syringe. When L is injected into P and the complex MP is formed, heat is produced or absorbed. Thus, the sample cell solution heats up or cools down, and $\Delta T \neq 0$. The power needed to equilibrate the system to $\Delta T \approx 0$ again is the signal detected over time in an ITC experiment.

The quantitative thermodynamic characterization of macromolecular interactions renders ITC a powerful method, especially since it has no molecular weight limitation and does not require labeling with fluorophors.

An ITC machine (Figure 21) consists of a reference and a sample cell with volume V_0, both enclosed in an adiabatic jacket. The cells are constantly heated, and the temperature difference between them is very small, but can be measured. The macromolecule solution (M) with concentration M_0 is filled into the sample cell, while the ligand solution (L) with concentration L_0 is filled into the stirring syringe. During the measurement, a defined volume V_{inj} of ligand solution is injected into the sample cell at predefined time points. The injection changes the concentrations of protein and ligand to $M_{tot,i}$ and L_i, respectively. Upon interaction of the protein and ligand, the complex formation induces heat release or absorbance from the sample cell solution. Thus, the temperature of the sample cell is altered. The time-dependent power difference Q_i that is needed to keep ΔT close to zero is measured in the ITC experiment and defined

$$\boldsymbol{Q_i = \langle \Delta H \rangle_i M_{tot,i} V_0 - \langle \Delta H \rangle_{i-1} M_{tot,i-1} (V_0 - V_{inj})} \qquad (29)$$

where $<\Delta H>_i$ is the change in enthalpy per mole of protein, $M_{tot,i}$ is the concentration of protein after injection i, V_0 is the cell volume, $M_{tot,i-1}$ is the concentration of protein before injection i and V_{inj} is the injection volume. In case of a stoichiometry of $N = 1$,

$$\boldsymbol{\langle \Delta H \rangle_i = \Delta H \frac{ML_i}{M_{tot,i}} = \Delta H \frac{ML_i}{M + ML_i} = \Delta H \frac{K_a L_i}{1 + K_a L_i}} \qquad (30)$$

where ΔH is the change in enthalpy when one mole of ligand binds to one mole of protein, L_i is the ligand concentration after injection i and K_a is the dissociation constant. Equations 30 and 31 can be combined and yield

$$\boldsymbol{Q_i = \frac{\Delta H M_{tot,i} V_0 K_a L_i}{1 + K_a L_i} - \frac{\Delta H M_{tot,i-1} (V_0 - V_{inj}) K_a L_{i-1}}{1 + K_a L_{i-1}}} \qquad (31)$$

The concentrations of protein $M_{tot,i}$ and ligand L_i can always be calculated from the cell and injection volumes V_0 and V_{inj} and the initial concentrations M_0 and L_0, and Q_i is the parameter measured in the experiment. Hence, fitting of the experimental ITC curve to equation 31 results in the thermodynamic parameters of the binding interaction, meaning the binding constant K_a, the stoichiometry N and the enthalpy ΔH. The entropy ΔS can then be derived using

$$\boldsymbol{\Delta G = -RT \ln K_a = \Delta H - T \Delta S} \qquad (32)$$

In order to conduct a successful ITC experiment, it is important to choose the protein concentration according to an appropriate C-value

$$\boldsymbol{C} = [\boldsymbol{M_0}]/\boldsymbol{K_D} \tag{33}$$

where [M] is the concentration of protein and K_D is the dissociation constant which is equal to $1/K_a$. The optimum C-value ranges between 10 and 100, for instance if the K_D is 100 nM, a protein solution of 5 µM concentration should be used in the experiment. For N = 1, the ligand concentration should be 10-fold the protein concentration (50 µM in this case). If the C-value become too low or too high, curve fitting will be more difficult or even impossible.

5.6.2 Isothermal Titration Calorimetry – Experiments and Data Evaluation

ITC experiments were performed using a Nano ITC Low Volume (TA Instruments) or a MicroCal™ VP-ITC System (GE Healthcare). Proteins were prepared in SEC buffer at a concentration of 100 µM (Nano ITC Low Volume) or 20 µM (MicroCal™ VP-ITC). Ligands were prepared in the same buffer at concentrations of 1 mM (Nano ITC Low Volume) or 200 µM (MicroCal™ VP-ITC). If concentrations differed from these values, it will be indicated. Titration experiments were performed at 307 rpm stirring speed and 25 °C. The ligand solution was injected into the sample cell at 0.5 µL/s and with 200-250 s gap time between single injections. Data were visualized using *NanoAnalyze* (TA Instruments), *Origin7* and *Origin9* (OriginLab, Northampton, MA). The programs *NITPIC*, *Sedpha*t and *Gussi* were utilized for thorough data analysis and figure preparation (Houtman *et al.*, 2007; Keller *et al.*, 2012).

5.7 Microbiological Methods

5.7.1 Preparation of M9 Agar for Spot Plate Assays

M9 glucose minimal agar (M9) plates for spot plate assays were prepared by heating 63.5 mL of 2x agar (30 g/l) to boiling in a microwave. 25 mL of 5x M9 salts, 34.5 mL of ddH_2O and 3 mL of glucose-salt stock were added and mixed by shaking. 45 mL of this solution were transferred to a sterile falcon tube containing an appropriate amount of ligand stock solution. The agar solution was mixed and 15 mL were poured into a petri dish for cooling. Agar plates were freshly prepared on the day of the experiment. Pyocyanin and bleomycin were used in spot plate assays to investigate the physiological functions of PA0803, PA1353, PA3127, PA3128 and PA4641. A 50 mM pyocyanin stock solution in 100% DMSO was used to add pyocyanin to the M9 agar plates in final concentrations 10 µM, 100 µM and 1 mM. An appropriate amount of DMSO was added to keep its concentration at the same level (0.02%) in all plates. A stock solution of 1 mM bleomycin in SEC buffer was used to add bleomycin added to the plates in final concentrations of 1 µg/mL, 0.5 µg/mL as well as 0.1 µg/mL.

5.7.2 Spot Plate Assay in *E. coli*

Spot plate assays were used to demonstrate the function of βαβββ-module proteins in living cells. *E. coli* BL21(DE3) cells were transformed with the respective plasmids according to 5.2.13 and grown on LB agar supplemented with appropriate antibiotics at 37 °C overnight. A 4 mL culture supplemented with appropriate antibiotics was inoculated with a single colony and grown again at 37 °C overnight. 200 µL of this culture were used to inoculate a 50 mL LB culture supplemented with appropriate antibiotics. The 50 mL culture was grown to an OD_{600} of 0.8 and induced with 0.5 mM IPTG. After overnight protein production at 20 °C under vigorous shaking, the OD_{600} of the cells was normalized to 1 and a serial dilution was prepared in fresh LB medium without antibiotics in a sterile 96-well plate. Using a multichannel pipette, 5 µL of the appropriate dilutions were spotted on agar plates. The spots were allowed to dry under the clean bench. Plates were incubated at 37 °C overnight and imaged on the following day.

5.7.3 Spot Plate Assay in *P. aeruginosa*

5.7.3.1 Transposon Mutants

P. aeruginosa $PAO1_{wt}$ and $PA14_{wt}$, as well as transposon mutants ΔPA0803, ΔPA1353, ΔPA3127 and ΔPA4641 (Table 3, Liberati *et al.*, 2006; Held *et al.*, 2012) were cultivated in LB medium overnight at 37 °C to regenerate them from glycerol shock. On the following day, a fresh culture of LB medium was inoculated with 5 µL from the overnight culture and again grown at 37 °C ON using appropriate antibiotics. On the next day, the OD_{600} of the cells was normalized. Using a multichannel pipette, a serial dilution was prepared in fresh LB medium without antibiotics in a sterile 96-well plate. 5 µL of the appropriate dilutions were spotted on the freshly prepared agar plates (Section 5.7.1). The spots were allowed to dry under the clean bench. Plates were incubated at 37 °C overnight and imaged on the following day.

5.7.3.2 Gain-of-Function Strains

P. aeruginosa $PAO1_{wt}$ and *P. aeruginosa* gain-of-function strains (Table 3) were prepared using *P. aeruginosa* PAO1 magnesium-competent cells (Section 5.2.12) which were transformed (Section 5.2.13) with pHERD30T plasmids (Table 2), each carrying one gene encoding a βαβββ-module protein PA0803, PA1353, PA3127 or PA4641. The gain-of-function strains were cultivated in LB medium overnight at 37 °C to regenerate them from glycerol shock. On the following day, a fresh culture of LB medium supplemented with 12 µg/mL gentamicin was inoculated with 5 µL from the overnight culture and again grown at 37 °C. At an $OD_{600} = 0.8$, the cells were induced with 0.1% arabinose and grown at 20 °C overnight. On the next day, the OD_{600} of the cells was normalized and a serial dilution was prepared in fresh LB medium without antibiotics in a sterile 96-well plate. Using a multichannel pipette, 5 µL of the appropriate dilutions were spotted on the freshly prepared agar plates (Section 5.7.1). The spots were allowed to dry under the clean bench. Plates were incubated at 37 °C overnight and imaged on the following day.

6 Results and Discussion

This work deals with βαβββ-module containing proteins from *P. aeruginosa* PAO1. First, general results on the identification and classification of proteins of this fold family from *P. aeruginosa* will be discussed. Then, a selection of βαβββ-module containing proteins from *P. aeruginosa* PAO1 which were examined during this study will be discussed. For this purpose, the proteins will be split into functional groups and discussed separately.

6.1 General Aspects of the βαβββ-Module Protein in *P. aeruginosa* PAO1

6.1.1 Identification of βαβββ-Module Proteins in *P. aeruginosa* PAO1

In earlier studies, three EhpR-like βαβββ-module containing proteins from *P. aeruginosa* PAO1 have been identified: PA0803, PA1353 and PA4641. These proteins have been characterized biophysically by ITC and structurally by X-ray crystallography. They bind the blue phenazine pyocyanin, an important *P. aeruginosa* virulence factor, with low µM K_D. In order to identify further βαβββ-module containing proteins from *P. aeruginosa* PAO1, the protein fold recognition server *Phyre²* (Protein Homology/analogy Recognition Engine V 2.0) was employed in a *BackPhyre* search (Kelley & Sternberg, 2009). In this *BackPhyre* search, a known protein structure is used to browse the genome of a target organism to identify genes that might produce proteins of similar fold. In the *BackPhyre* procedure, the sequence and secondary structure information is extracted from the 3D structure provided by the user to build a Hidden Markov model (HMM). This model is matched to the HMM database of all proteins encoded in the genome of interest.

Using the crystal structure of EhpR from *P. agglomerans* (Yu *et al.*, 2011), the genome of *P. aeruginosa* PAO1 was searched for additional βαβββ-module containing proteins. In this *BackPhyre* search the *Phyre²* server returned a list of 21 βαβββ-module containing proteins. The *BackPhyre* search was repeated using the crystal structures of PA0803 (Yu, 2009), PA1353 and PA4641 (Kalawy-Fansa, 2010). Again, the search algorithm yielded the same 21 βαβββ-module proteins found using EhpR as a search model, but in slightly different order (Table 18).

Table 18: Result of the *BackPhyre* search with EhpR in the *P. aeruginosa* PAO1 genome that identified 21 probable βαβββ-module containing proteins. PA4641 was not identified by *BackPhyre* because it was thought to be a pseudogene. In total, 22 βαβββ-module containing proteins are present in the *P. aeruginosa* genome. The rank, residues modeled, the confidence of the model and the sequence identity (Id.) are indicated for the search with EhpR in *BackPhyre*. These are then compared to sequence identity (Id.) of the results from PA0803, PA1353 and PA4641. The results show the low sequence identity, but high structural similarity among the βαβββ-module containing proteins.

Rank EhpR	Protein	Total No. of Residues	Residues modeled	Conf. / %	Id. / %	Id. to PA0803 / %	Id. to PA1353 / %	Id. to PA4641 / %
1	PA3956	144	1-132	99.8	16	20	16	17
2	PA0817	141	3-133	99.8	17	20	13	11
3	PA3389	137	1-127	99.7	14	16	17	14
4	PA4518	141	5-128	99.7	20	24	23	16
5	PA4183	155	31-148	99.7	11	23	15	17
6	PA0880	126	2-125	99.7	13	17	9	15
7	PA1129	135	1-115	99.7	16	19	21	14
8	PA0803	146	1-143	99.7	16	100	14	9
9	PA2546	143	1-137	99.7	16	16	16	14
10	PA1672	126	1-121	99.7	15	23	17	11
11	PA2024	140	4-126	99.7	11	22	12	11
12	PA5111	176	15-168	99.7	17	16	15	12
13	PA3503	211	36-157	99.6	11	17	13	13
14	PA3524	128	1-126	99.6	14	15	23	15
15	PA0710	131	1-127	99.6	12	19	13	18
16	PA3127	267	1-115	99.6	18	28	24	19
17	PA1353	137	4-133	99.3	12	14	100	44
18	PA1358	132	1-131	99.1	15	16	25	29
19	PA0865	357	5-138	98.9	5	13	11	13
20	PA2721	159	2-121	94.9	13	20	21	20
21	PA0242	634	287-413	93.4	6	18	8	8
(22)	PA4641	137	-	-	-			100

In total, the genome of *P. aeruginosa* PAO1 contains 22 βαβββ-module proteins. The secondary structure elements of these 22 proteins were modeled with high confidence (> 90%), even if most of them share relatively low sequence identity (< 30%). However, 22 similar proteins are much more than expected. According to *SCOPe* Database (Version 2.04 2014-09-11), there are 1205 different protein folds known (Fox *et al.*, 2014). Since *P. aeruginosa* PAO1 possesses 5570 genes (Stover *et al.*, 2000), statistically every protein fold should occur between four and five times, but since the βαβββ-module fold is present in 22 proteins of *P. aeruginosa* PAO1, this fold seems to be privileged. Possibly, this is because the modules can be combined easily by gene duplication and horizontal gene transfer. Since the binding or active sites are located at the dimer interfaces, few mutations may be sufficient to generate new functions. The low sequence similarity also indicates that mutations might have been introduced all over the protein, as long as the 3D structure and fold were conserved. Therefore, only very few residues are conserved in the sequence. Due to the low sequence identity among the βαβββ-module proteins, it is impossible to assign functions any of the 22 βαβββ-module proteins in *P. aeruginosa* PAO1 just by bioinformatics approaches. Nevertheless, it may be possible to describe some general characteristics of these proteins and then to classify them into subgroups.

6.1.2 Functionally or Structurally Characterized βαβββ-Module Proteins

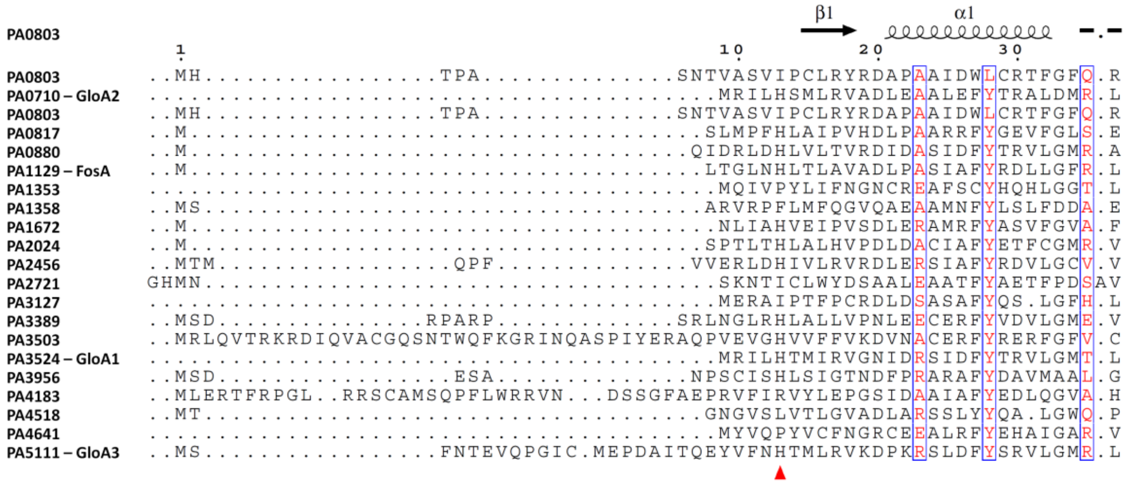
Four of the 22 βαβββ-module proteins of *P. aeruginosa* PAO1 have been functionally characterized. The proteins PA0701, PA3524 and PA5111 were found to detoxify methylglyoxal. Thus, PA0701, PA3524 and PA5111 were termed GloA2, GloA1 and GloA3, respectively. GloA enzymes have been described in pro- and eukaryotes (Sukdeo & Honek, 2007). Interestingly, two biochemically distinct classes of GloA enzymes, a Zn(II)-activated and a non-Zn(II)-activated class exist (Sukdeo *et al.*, 2004). Surprisingly, *P. aeruginosa* possesses enzymes from both metal activation classes: GloA3 belongs to the Zn(II)-containing holoenzymes, while GloA1 and GloA2 can be activated by either Ni(II) or Co(II). Furthermore, *P. aeruginosa* is the only organism known to date to possess three of these enzymes (Sukdeo & Honek, 2007). In contrast, PA1129 (PDB entry 1NNR) was characterized as a fosfomycin resistance protein FosA *in vitro* and *in vivo* (Rife *et al.*, 2002; Rigsby *et al.*, 2004). This Mn(II)- and K^+-dependent enzyme catalyzes the detoxification of fosfomycin under GSH consumption.

Structural analysis of five of the 22 βαβββ-module proteins of *P. aeruginosa* PAO1 has been performed in the past decades. The crystal structures of PA2721 (PDB entry 1U69, (Nocek *et al.*, 2006) and PA1358 (PDB entry 1U7I, Osipiuk *et al.*, to be published) have been determined. However, both proteins have not been examined biochemically and there is no hint for a function of these proteins. Furthermore, the crystal structures of

PA0803, PA1353 and PA4641 have been determined and a preliminary function as pyocyanin resistance proteins has been assigned to them using ITC and soaking techniques (Yu, 2009; Kalawy-Fansa, 2010).

6.1.3 Sequence Alignment of *P. aeruginosa* PAO1 βαβββ-Module Proteins

In order to classify the βαβββ-module proteins of *P. aeruginosa* PAO1, the structural alignment program *PROMALS3D* (Pei *et al.*, 2008). The crystal structure of PA0803 was utilized together with the protein sequences of all βαβββ-module proteins comprising two modules per monomer (Figure 22 and Figure 23).



```
                  1                                       10        20        30
PA0803          ..MH.................TPA..............SNTVASVIPCLRYRDAPAAIDWLCRTFGFQ.R
PA0710 – GloA2  .........................................MRILHSMLRVADLEAALEFYTRALDMR.L
PA0803          ..MH.................TPA..............SNTVASVIPCLRYRDAPAAIDWLCRTFGFQ.R
PA0817          ..M.....................................SLMPFHLAIPVHDLPAARRFYGEVFGLS.E
PA0880          ..M....................................QIDRLDHLVLTVRDIDASIDFYTRVLGMR.A
PA1129 – FosA   ..M.....................................LTGLNHLTLAVADLPASIAFYRDLLGFR.L
PA1353          .........................................MQIVPYLIFNGNCREAFSCYHQHLGGT.L
PA1358          ..MS....................................ARVRPFLMFQGVQAEAAMNFYLSLFDDA.E
PA1672          ..M......................................NLIAHVEIPVSDLERAMRFYASVFGVA.F
PA2024          ..M.....................................SPTLTHLALHVPDLDACIAFYETFCGMR.V
PA2456          ..MTM.................QPF..............VVERLDHIVLRVRDLERSIAFYRDVLGCV.V
PA2721          GHMN.....................................SKNTICLWYDSAALEAATFYAETFPDSAV
PA3127          .........................................MERAIPTFPCRDLDSASAFYQS.LGFH.L
PA3389          ..MSD...............RPARP.............SRLNGLRHLALLVPNLEECERFYVDVLGME.V
PA3503          ..MRLQVTRKRDIQVACGQSNTWQFKGRINQASPIYERAQPVEVGHVVFFVKDVNACERFYRERFGFV.C
PA3524 – GloA1  .........................................MRILHTMIRVGNIDRSIDFYTRVLGMT.L
PA3956          ..MSD.................ESA..............NPSCISHLSIGTNDFPRARAFYDAVMAAL.G
PA4183          ..MLERTFRPGL..RRSCAMSQPFLWRRVN...DSSGFAEPRVFIRVYLEPGSIDAAIAFYEDLQGVA.H
PA4518          ..MT....................................GNGVSLVTLGVADLARSSLYYQA.LGWQ.P
PA4641          .........................................MYVQPYVCFNGRCEEALRFYEHAIGAR.V
PA5111 – GloA3  ..MS.................FNTEVQPGIC.MEPDAITQEYVFNHTMLRVKDPKRSLDFYSRVLGMR.L
```

Figure 22: Part I of the structural alignment of βαβββ-module proteins with *PROMALS3D* (Pei *et al.*, 2008). Sequences of all proteins containing two modules per monomer were aligned to the three-dimensional crystal structure of PA0803. Red triangles indicate residues conserved for metal-binding (H, H, and E), blue stars indicate residues conserved for aromatic compound binding (two out of F, H, W, and Y).

The 22 βαβββ-module proteins are not very homologous regarding their overall sequence. Still, some amino acids are abundant in these proteins, for instance alanine, aspartate, glycine and leucine. Furthermore, there are several positions in which amino acids are structurally conserved. Some proteins possess two conserved histidine residues and a glutamate in the active center (red triangle). In others these residues are not conserved, but instead they include two aromatic residues that could be phenylalanine, histidine, tryptophan or tyrosine (blue star). In all remaining proteins, either not all or none of these residues important for metal- or aromatic compound-binding are conserved.

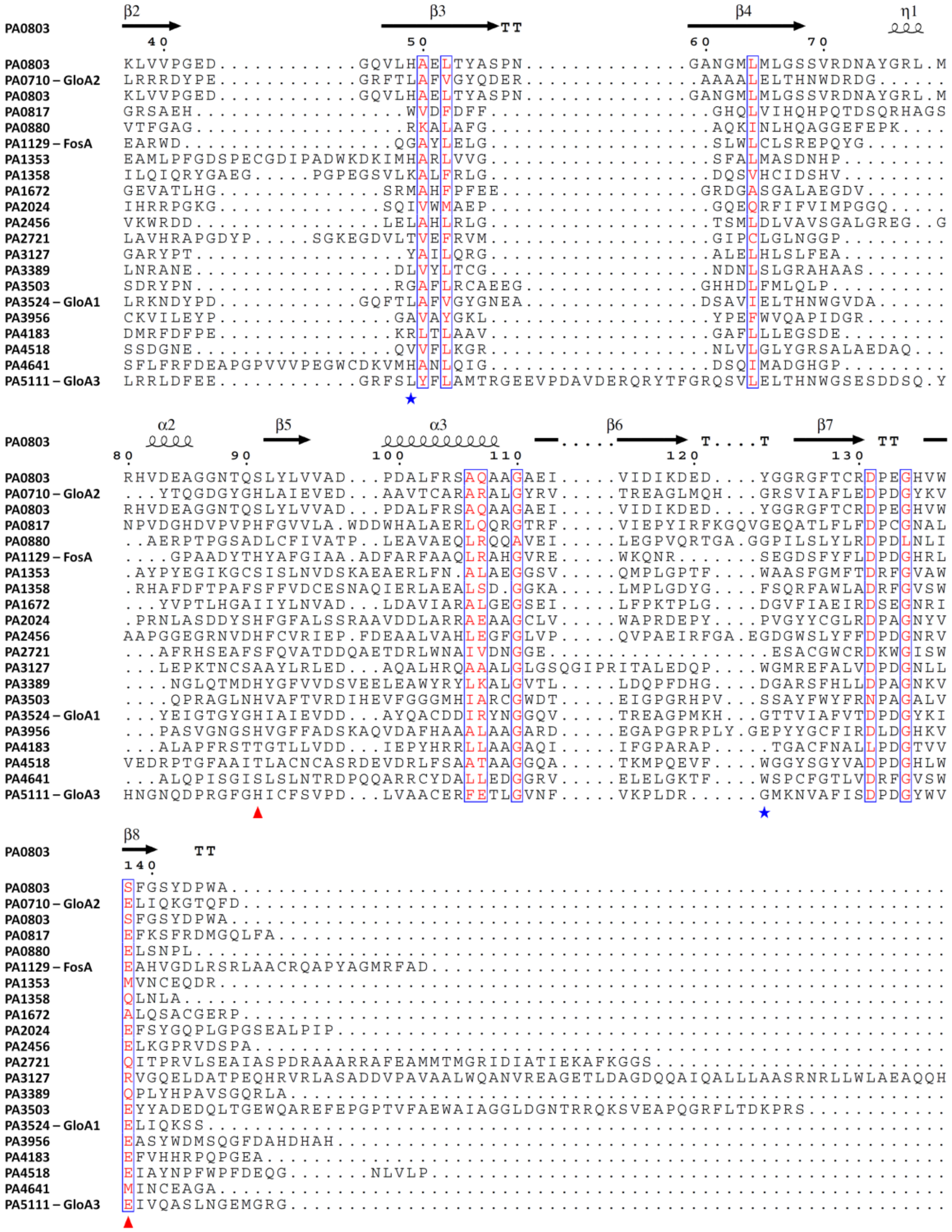


Figure 23: Part II of the structural alignment of βαβββ-module proteins with *PROMALS3D* (Pei *et al.*, 2008). The sequences of all proteins containing two modules per monomer were aligned to the three-dimensional structure of PA0803. Red triangles indicate residues conserved for metal-binding (H, H, and E), blue stars indicate residues conserved for aromatic compound binding (two out of F, H, W, and Y).

Further residues are conserved in few positions in the proteins. The residues probably are essential to maintain the βαβββ-module structure or the hydrophobic interface at the core of the proteins. An example of conserved residues is given in Figure 24.

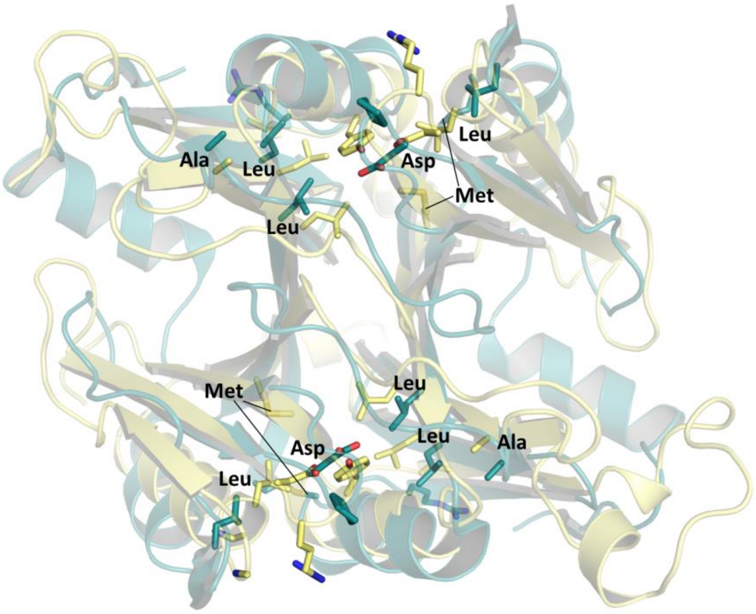


Figure 24: Structurally conserved residues (colored in red in the alignment, Figures 22 and 23) in βαβββ-module resistance proteins from *P. aeruginosa*, using PA1353 (yellow) and FosA (PA1129, teal) as examples. Conserved residues are shown as sticks. They are mainly located at the interface of two modules within one monomer. Hence, they stabilize the three-dimensional structure of the βαβββ-module resistance proteins.

PA1129 (FosA) and PA1353 were chosen as examples to analyze the residues in structurally conserved positions. These residues are preferentially found at the intramolecular interface of the of βαβββ-module protein monomers. Most of these residues are hydrophobic, for instance leucine, alanine or methionine, but also an aspartate is conserved at this interface. These residues are most likely important to stabilize the three-dimensional structure of the βαβββ-module proteins. In contrast, few conserved residues are found in the binding site, as mentioned above.

6.1.4 Classification of βαβββ-Module Proteins from *P. aeruginosa* PAO1

On the basis of the structural alignment with *PROMALS3D*, it is possible to classify the βαβββ-module proteins from *P. aeruginosa* PAO1 into three classes I, II and III based on the conserved residues in the active center (Table 19).

In class I, the metal-binding βαβββ-module proteins can be found. These proteins possess two conserved histidine residues and a glutamate in the active center. They coordinate a divalent metal cation together with a water molecule. Possible cations are Mn(II), Fe(II), Co(II), Ni(II), Zn(II) or Mg^{2+} (Armstrong, 2000), depending on the reaction catalyzed by the proteins. The class I βαβββ-module proteins are subgrouped according to the number of βαβββ-modules per monomer. In subgroup I, all proteins with two modules per monomer can be found. In subgroup II, the proteins with 4 modules are located. All class I proteins probably catalyze detoxification reactions, such as the degradation of methylglyoxal (PA0710, PA3524, PA5111) or fosfomycin (PA1129).

Table 19: Classification of βαβββ-module containing proteins in *P. aeruginosa* PAO1 according to alignment with *PROMALS3D* (Pei *et al.*, 2008). A $^{\#}$ indicates that the structure is already known, a $^{+}$ indicates an additional domain. Conserved residues are given in the second line.

Class I		**Class II**	**Class III**
Metal-Binding Proteins		Aromatic Compound Binding Proteins	Proteins of Unknown Function
H, H, E	H, H, E	Two out of F, H, W or Y	?
Subgroup I (2 modules)	Subgroup II (4 modules)		
PA0710 (GloA2)	PA0242^{+}	PA0803$^{\#}$	PA0880
PA0817	PA0856	PA1353$^{\#}$	PA1672
PA1129 (FosA)$^{\#}$		PA3127^{+}	PA2721$^{\#}$
PA2024		PA4641$^{\#}$	PA1358$^{\#}$
PA2546			PA4183
PA3389			PA4518
PA3503^{+}			
PA3524 (GloA1)			
PA3956			
PA5111 (GloA3)			

Class II βαβββ-module proteins are proteins that possess two aromatic residues in their binding pocket. These residues could be phenylalanine, histidine, tryptophan or tyrosine. The aromatic compound is supposed to be sandwiched between the two aromatic residues via π-stacking interactions. In contrast to class I proteins, class II βαβββ-module proteins do not alter their ligand chemically, but may act as a chaperone for the ligand.

Class III βαβββ-module proteins do neither contain all three metal-binding residues nor both aromatic compound binding residues conserved in their active center. Hence no functional prediction on these proteins is possible. The proteins might split up into more distinct classes if they are investigated structurally and biochemically.

6.1.5 βαβββ-Module Proteins in the Focus of This Study

In order to shed light on some of these βαβββ-module proteins from *P. aeruginosa* PAO1, this research focuses on the structural and functional characterization of βαβββ-module proteins from class II (aromatic compound binding proteins) and class III (proteins of unknown function). On the one hand, the potential pyocyanin resistance factors PA0803, PA1353 and PA4641 as well as the uncharacterized fourth class II protein PA3127 will be examined functionally and structurally. This approach will include investigation of PA1354 which is encoded in an operon together with PA1353 (Figure 25).On the other hand, the class III proteins of unknown function PA0880, PA1672, PA4183 and PA4518 will be subject to 3D structure determination.

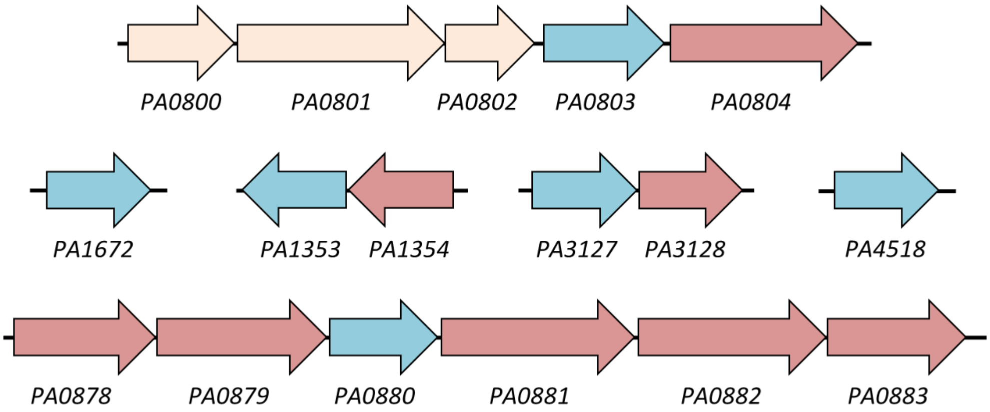


Figure 25: βαβββ-module proteins in the focus of the study and their predicted operons from *P. aeruginosa* PAO1 according to *DOOR* (Mao *et al.*, 2009). Fold, function or subcellular localization are indicated by the following color code: blue: βαβββ-module proteins; red: cytoplasm; light orange; cytoplasmic membrane.

6.2 Purification of Recombinant βαβββ-Module Proteins

The proteins investigated in this study were purified following a common protocol. Hence, only the purification of PA3127 is exemplarily described in detail here.

6.2.1 PA3127 as an Example Purification

The 267-amino acid protein PA3127 was expressed as an N-terminal His_6-tag fusion protein including a TEV protease cleavage site. The purification of PA3127 was performed using Ni(II)-IMAC, TEV protease cleavage during dialysis in SEC buffer and revNi(II)-IMAC.

As the final purification step, size exclusion chromatography was used to isolate a single oligomerization species of the protein (Figure 26).

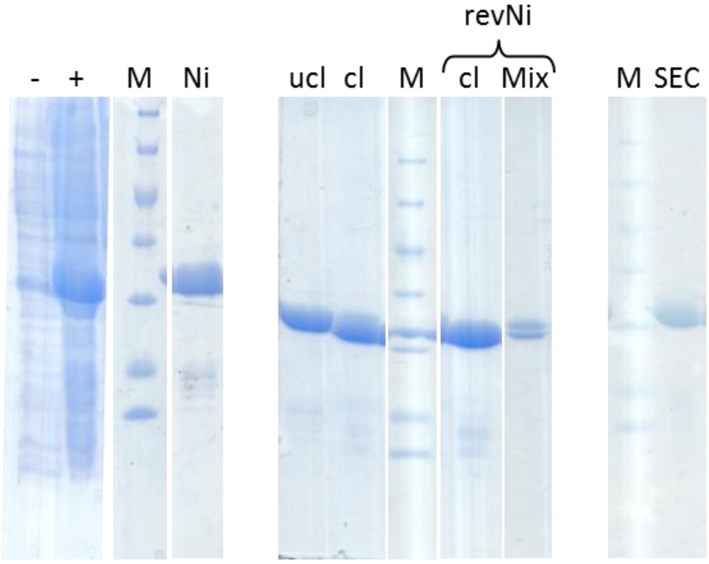


Figure 26: Purification of PA3127. PA3127 was expressed as a fusion protein with an N-terminal His$_6$-tag and a TEV protease cleavage site. The protein was purified via Ni(II)-IMAC and size exclusion chromatography. M: protein standard; -: before induction; +: after induction with IPTG; Ni: after Ni(II)-IMAC; ucl: uncleaved protein in dialysis; cl: cleaved protein (+TEV) in dialysis; revNi-cl: cleaved protein after revNi(II)-IMAC; Mix: mix of uncleaved and cleaved protein in revNi(II)-IMAC; SEC: PA3127 after size exclusion chromatography.

The purification yielded 8 mg of protein/L of culture. The protein was >90% pure as estimated from SDS-PAGE.

6.2.2 Final Purity of All βαβββ-Module Proteins

All proteins used in this study were produced and purified following a common protocol. Protein purification proceeded using Ni(II)-IMAC and size exclusion chromatography. If the His$_6$-tag was cleaved for crystallization, TEV protease was added to the dialysis and an additional Ni(II)-IMAC step was performed, before SEC was performed as the final purification step. Purification of His$_6$-PA0803, His$_6$-PA1353 and His$_6$-PA4641 has been described in detail before (Yu, 2009; Kalawy-Fansa, 2010). Additionally, purification of PA1672 and PA3127 has been described in studies under the supervision of the author (Weber, 2011; Popp, 2013). In this study, purification of PA0880, His$_6$-PA1354, PA4183_16-155, PA4183_32-155 and PA4518 have been performed for the first time. The final purity of each protein purified in this study is shown in Figure 27.

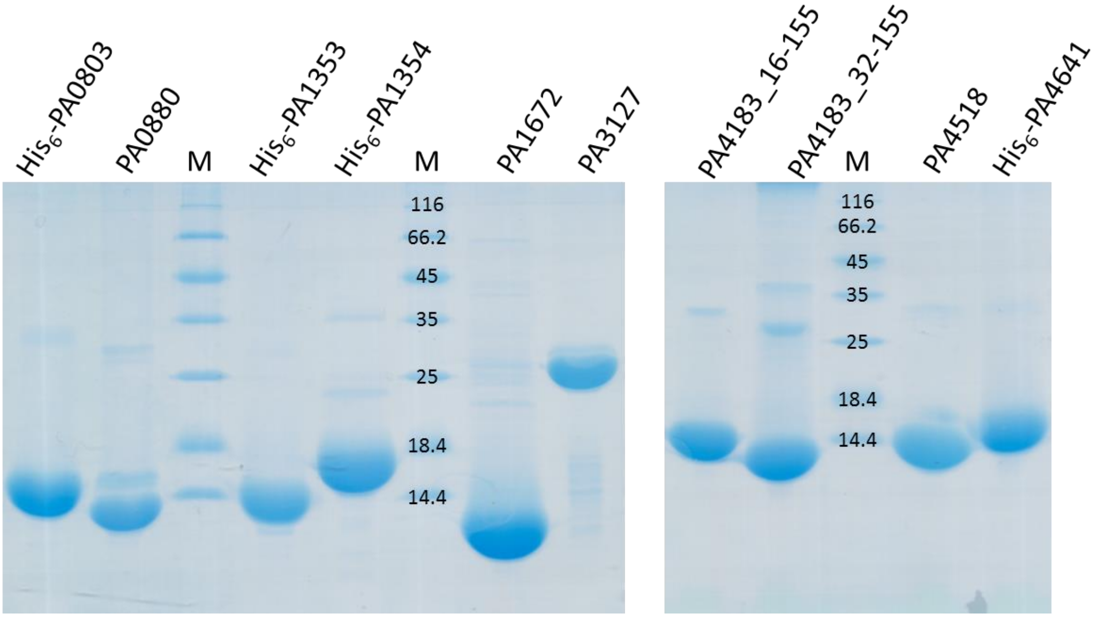


Figure 27: Final purity of the proteins after the final SEC purification step. 20 µg of protein were loaded and compared to a molecular weight standard. This gel is used to estimate the final purity of the protein. Most proteins have the expected molecular weight (Table 20). PA1672 is running at a lower molecular weight than expected. Most proteins are highly pure >90% as estimated).

All proteins were isolated in high purity and yield from *E. coli*. A summary on the protein purification of the proteins investigated in this study, including yield and purity estimations, is given in Table 20.

Table 20: Protein purification summary. The type of protein, the molecular weight, the purification strategy (2 = IMAC, SEC; 4 = IMAC, TEV cleavage, revIMAC, SEC) and the yield are listed in this table. Purities are estimated from SDS-PAGE gels. - = not subjected to crystallization.

Protein	MW / kDa	Purification Steps	Oligomer in SEC	Yield / mg/L of culture	Final Purity / %	Crystals
His$_6$-PA0803	17.94	2	dimer	50	>95	yes
PA0880	13.87	4	dimer	22	>85	yes
His$_6$-PA1353	15.28	2	dimer	50	>95	-
His$_6$-PA1354	18.03	2	dimer	31	>90	yes
PA1672	13.61	4	dimer	5	>90	yes
PA3127	29.60	4	dimer	8	>90	yes
PA4183_16-155	15.58	4	dimer	35	>90	yes
PA4183_32-155	13.80	4	dimer	32	>80	no
PA4518	15.35	4	dimer	20	>95	yes
His$_6$-PA4641	17.28	2	dimer	68	>95	-

6.3 Proteins of Unknown Function – PA1354, PA4183, PA1672 and PA4518

6.3.1 PA1354

6.3.1.1 Bioinformatic Analysis of PA1354

The cytoplasmic protein PA1354 is predicted to be encoded in an operon together with PA1353, a probable pyocyanin resistance factor. In this study, PA1354 was produced as His_6-PA1354 fusion protein. This construct yields a 158-amino acid protein with a calculated molecular weight of 18.03 kDa and a theoretical pI of 5.12 (Gasteiger *et al.*, 2003). The amino acid sequence of PA1354 was submitted the PDB, which returned only one protein structure, the hypothetical protein PA1349 from *P. aeruginosa*, with a sequence identity of 26%. Additionally, the sequence of PA1354 was submitted to the structure prediction server *Phyre²* (Table 21), in order to identify structurally related proteins and a possible function. The model confidence for PA1354 significantly dropped after the first three hits.

Table 21: Models used for *Phyre²* predictions of PA1354. PF03795 = YCII related domain, PF02426 = muconolactone δ-isomerase.

PDB ID	Functional Annotation	Sequence Identity / %	Model Confidence / %	PFAM family
1S7I	PA1349 from *P. aeruginosa*; ferredoxin-like, YCII-related domain	29	100	PF03795
4LBH	5-chloro-2-hydroxyhydroquinone dehydrochlorinase from *Burkholderia phenoliruptrix*	24	98.2	PF03795
1MWQ	hypothetical protein from *Haemophilus influenzae* with a putative active-site phosphohistidine	13	98.1	PF03795
3ZNU	chloromuconolactone dehalogenase ClcF from *Rhodococcus opacus*	20	89.8	PF02426
1MLI	muconolactone isomerase from *Pseudomonas putida*	19	79.7	PF02426

On the one hand, *Phyre²* relates PA1354 to muconolactone δ-isomerases (PF02426) that catalyze the third step in the conversion of catechol to succinate and AcCoA. On the other hand, *Phyre*^{2} predicts PA1354 to be a ferredoxin-like, YCII-related domain protein (PF03795). According to the *Protein Families Database* PFAM (Finn *et al.*, 2014), this domain might have an enzymatic function catalyzed by a conserved histidine and an aspartate residue. *Caulobacter crescentus* possesses a protein (*UniProt* entry Q9A8M4) where a YCII domain is fused to a σ70 factor, suggesting that it might play a role in transcription initiation. Thus, PA1354 possibly interacts with a σ-factor to initiate the expression of distinct genes in *P. aeruginosa*.

6.3.1.2 Crystallization of His$_6$-PA1354

Three initial crystallization hits were identified for His$_6$-PA1354 (JCSG Core I G2, H1 and JCSG Core II F7). All of them yielded plate-shaped crystals. After optimization, well-diffracting crystals were obtained from 0.1 M sodium acetate pH 5, 20% (*w/v*) MPD (JCSG Core I G2) (Figure 28). In diffraction experiments, they diffracted to 2.5 Å.

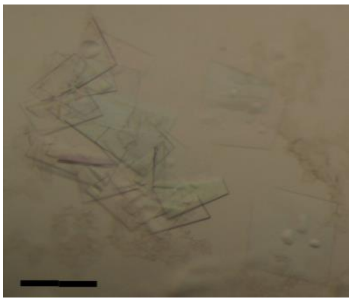


Figure 28: Crystals of His$_6$-PA1354 are fragile plates, but diffracted to 2.5 Å. Scale bar is 200 µm.

6.3.1.3 Data Collection and Structure Solution Attempts of His$_6$-PA1354 Crystals

Optimized crystals of His$_6$-PA1354 were flash cooled in liquid nitrogen without additional cryoprotectants and subjected to data collection on beamline 14.1 at BESSY II Synchrotron of the HZB, Berlin, Germany. 105 images of 1° were collected on a MarCCD Chess detector using a wavelength of $\lambda = 0.9184$ Å. Data were integrated with *XDS* (Kabsch, 2010) and scaled with *Aimless* from the *CCP4 program suite* (Evans, 2006, 2011; Winn *et al.*, 2011). The space group of His$_6$-PA1354 crystals could not be determined unambiguously, but is most probably $P2_12_12$ as analyzed by *POINTLESS* from the CCP4 suite (Winn *et al.*, 2011). Data collection and processing statistics are given in Table 22.

Table 22: Data collection statistics of a *P. aeruginosa* PAO1 His$_6$-PA1354 crystal. Values in parentheses are for the highest resolution shell.

Dataset	His$_6$-PA1354
Wavelength (Å) / beamline[≠]	0.9184 / BESSY II, 14.1
Resolution range (Å)	52.04-2.5 (2.6-2.5)
Space group	$P2_12_12$
Unit cell parameters (Å) (°)	52.04 61.46 50.36 90 90 90
Mosaicity (°)[†]	0.192
Total No. of measured reflections	24521 (2749)
Unique reflections	5927 (647)
Multiplicity	4.1 (4.2)
Mean I/σ(I)	14.4 (2.0)
Completeness (%)	99.7 (99.8)
R_{meas} (%)	6.8 (53.5)
$R_{p.i.m.}$ (%)	3.3 (39.1)

[†] Mosaicity values reported by *XDS* (Kabsch, 2010)

Various MR attempts to determine the structure of His$_6$-PA1354 failed. There are only few models available in the PDB, as predicted and seen in the previous *BLAST* and *Phyre*^{2} searches (Table 22). All of them were used for MR attempts, as well as the MR engine *BALBES*. Thus, phasing strategies such as selenomethionine labeling or heavy atom derivatization will have to be applied in the future in order to phase the structure of PA1354.

6.3.2 Class III βαβββ-Module Protein PA4183

6.3.2.1 Bioinformatic Analysis of PA4183

The sequence of the cytoplasmic protein PA4183 was available from the *Pseudomonas Genome Database* (Winsor *et al.*, 2011) and it was identified as a class III βαβββ-module protein in *P. aeruginosa* PAO1 genome by the *BackPhyre* search using EhpR. According to the alignment using *PROMALS3D* (Figure 22, Figure 23), PA4183 belongs to the class III βαβββ-module proteins. As contradictory information was available on which was the starting methionine of the open reading frame of PA4183, three constructs His$_6$-PA4183, His$_6$-PA4183_16-155 and His$_6$-PA4183_32-155 were cloned. A *BLAST* Search in the PDB using the sequence of PA4183 returned one hit, an uncharacterized protein MM3218 from *Methanosarcina mazei* without any functional characterization. *Phyre^2* analysis of PA4183 returned a high confidence model, but no hint to a function for PA4183.

6.3.2.2 Crystallization of PA4183_16-155

While His$_6$-PA4183_32-155 was insoluble, His$_6$-PA4183 and His$_6$-PA4183_16-155 were soluble and could be purified. However, only His$_6$-PA4183_16-155 crystallized in two conditions, JCSG Core I B1 and D7, yielding long needles and rods, respectively. However, after optimization no diffracting crystals of PA4183_16-155 were obtained (Figure 29).

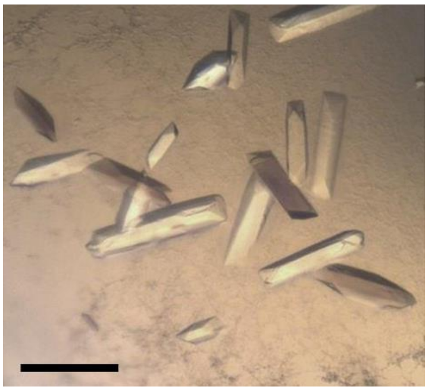


Figure 29: Rod shaped crystals of PA4183_16-155 showed no diffraction in X-ray test experiments. Scale bar is 200 µm.

6.3.3 Class III βαβββ-Module Protein PA1672

6.3.3.1 Bioinformatic Analysis of PA1672

PA1672 was produced as a His$_6$ fusion protein in *E. coli* and purified via Ni^{2+}-IMAC and TEV cleavage of the His$_6$ tag during dialysis overnight. Size exclusion chromatography was performed as the last purification step, yielding PA1672 with a calculated molecular weight of 13.61 kDa, respectively (Gasteiger *et al.*, 2003). PA1672 is a class III βαβββ-module resistance protein of unknown function, lacking two of the three metal-binding residues as well as both aromatic compound binding residues. The gene encoding PA1672 is predicted not to be included in operons (Mao *et al.*, 2009). Hence, the genetic context is probably not helpful to understand the function of PA1672.

A *BLAST* search in the PDB (E-value cut off 1.0) using the sequence of PA1672 only identified three hits with sequence similarity over the whole protein, namely RHA04853 from *Rhodococcus* sp. RHA1 (PDB entry 2RU6, 26% sequence identity), a glyoxalase I from *Listeria monocytoge*nes (PDB entry 3E5D, 26% sequence identity) and a glyoxalase/bleomycin resistance protein from *Albidiferax ferrireducens* (PDB entry 4QB5, 25% sequence identity). However, these three proteins lack functional characterization and do not provide a starting point for functional characterization of PA1672. *Phyre*^{2} structure predictions for PA1672 are accurate and predict the structure of PA1672 with high model confidence, and identified 120 potential templates for structural modeling. The top five hits are given in Table 23.

The proteins identified in the *Phyre*^{2} search are proteins from glyoxalase subfamilies I, II and IV, meaning from metal-binding enzymes or from antibiotic resistance proteins such as the mitomycin resistance proteins. For this reason, it is not possible to identify the function of PA1672 from this search.

Table 23: Models used for *Phyre*^{2} predictions of PA1672. PF00903 = glyoxalase I, PF12681 = glyoxalase II, PF13669 = glyoxalase IV. * = also identified in the *BLAST* search in the PDB.

PDB ID	Functional Annotation	Sequence Identity / %	Model Confidence / %	PFAM family
2R6U*	RHA04853 from *Rhodococcus* sp. RHA1	26	99.9	PF00903, PF12681
4G6X	glyoxalase/bleomycin resistance protein from *Catenulispora acidiphila*	23	99.8	PF00903, PF12681
2I7R	protein from *Streptococcus pneumoniae*	16	99.8	PF12681
1MPY	Catechol-2,3-dioxygenase (metapyrocatechase) from *P. putida*	23	99.8	PF00903
1QIP	Human glyoxalase I complexed with S-P-nitrobenzyloxycarbonylglutathione	16	99.8	PF00903, PF12681, PF13669

Conserved domain analysis of PA1672 with the Conserved Domain Database (CDD) (Marchler-Bauer *et al.*, 2015) correctly classifies PA1672 as a member of the glyoxalase superfamily, and identifies it as a catechol-dioxygenase, and more specifically as an SgaA_N-like protein. SgaA suppresses the growth disturbance caused by high osmolality and a high concentration of A-factor (2-isocapryloyl-3R-hydroxymethyl-γ-butyrolactone), during early growth in *Streptomyces griseus*. A-factor controls morphological differentiation and secondary metabolism in the *S. griseus* QS system. However, A-factor has not been found in Pseudomonads yet. The protein SgaA might also function as a streptomycin resistance protein. This indicates that is does not bind divalent metal cations, as predicted in the previous sequence alignment. However, the 3D structure of SgaA is not available and cannot be compared with PA1672.

6.3.3.2 Crystallization and Data Collection of PA1672

Two initial hits were identified for PA1672 (JCSG Core I H2 and Core IV D6). Both hits were optimized and yielded well diffracting crystals (Figure 30). However, crystals obtained in 4.3 M sodium chloride, 0.1 M HEPES/NaOH pH 7.5 (JCSG Core IV D6) showed perfect twinning. Thus, an untwinned crystal obtained in 0.1 M sodium acetate pH 4.5, 20% (*w/v*) MPD was used for crystallographic data collection.

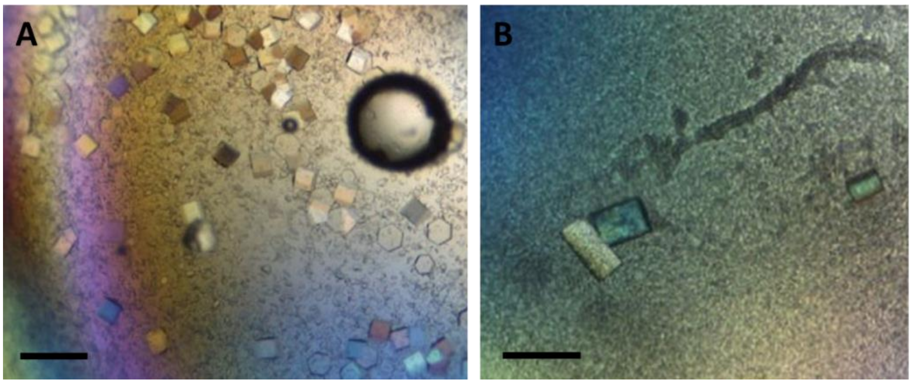


Figure 30: Crystals of PA1672. A: Perfectly twinned crystals in 4.3 M sodium chloride, 0.1 M HEPES/NaOH pH 7.5. B: Untwinned crystals in 0.1 M sodium acetate pH 4.5, 20% (*w/v*) MPD were used for data collection. Scale bar is 200 µm.

Crystals of PA1672 were cryoprotected in mother liquor supplemented with additional 5% (*w/v*) PEG 3350 / 15% (*v/v*) glycerol before flash cooling in liquid nitrogen. Diffraction data of PA1672 crystals were collected as 100 1° images on a MarCCD Chess detector at a wavelength of λ = 0.9184 Å on beamline 14.1 at BESSY II Synchrotron of the HZB, Berlin, Germany. Data collection and processing statistics are given in Table 24.

Table 24: Data collection and processing statistics for a *P. aeruginosa* PAO1 PA1672 crystal. Values in parentheses are for the highest resolution shell.

Dataset	PA1672
Wavelength (Å) / beamline$^{\neq}$	0.9184 / BESSY II, 14.1
Resolution range (Å)	44.13-1.81 (1.85-1.81)
Space group	$P2_12_12$
Unit cell parameters (Å) (°)	74.12 88.27 35.15 90 90 90
Mosaicity (°)†	0.292
Total No. of measured reflections	85590 (4348)
Unique reflections	21636 (1211)
Multiplicity	4.0 (3.6)
Mean I/σ(I)	17.6 (2.3)
Completeness (%)	99.4 (94.9)
R_{meas} (%)	5.4 (70.4)
$R_{p.i.m.}$ (%)	2.7 (36.3)

†Mosaicity values reported by *XDS* (Kabsch, 2010)

The PA1672 crystal diffracted to a resolution of 1.81 Å at the BESSY II Synchrotron and was indexed in orthorhombic space group $P2_12_12$ with cell parameters a = 74.12, b = 88.27 and c = 35.15. The Matthews coefficient (Matthews, 1968) was calculated as 2.11 Å^3/Da which corresponds to a solvent content of 42% and two PA1672 molecules (one dimer) in the asymmetric unit.

6.3.3.3 Structure Solution and Refinement of PA1672

The structure of PA1672 was determined by MR in *PHASER* from *CCP4* using a dimer of RHA04853 from *Rhodococcus* sp. RHA1 (PDB entry 2R6U, Zhang *et al.,* to be published) in *BALBES*. RHA04853 is an uncharacterized protein and has 28% sequence identity to PA1672. The model building and refinement yielded the structure of PA1672 with R_{work} = 19.8% and R_{free} = 24.1% and with 97.9% of the residues in Ramachandran-favored regions. The *MolProbity* score of the PA1672 structure is 1.20 (Table 25).

Table 25: Refinement statistics for *P. aeruginosa* PAO1 PA1672. Values in parentheses are for the highest resolution shell.

Dataset	PA1672
Resolution range (Å)	44.13-1.81 (1.85-1.81)
R_{work} (%)	19.8 (24.9)
R_{free} (%)	24.1 (29.7)
No. of non-H atoms	
Protein	1905
Ion	-
Ligand	-
Water	120
R.m.s. deviations	
Bonds (Å)	0.007
Angles (°)	1.013
Average B factors ($Å^2$)	
Protein	38
Ion	-
Ligand	-
Water	41
Ramachandran plot	
Favored regions (%)	97.9
Outliers (%)	0
MolProbity score[#]	1.20

[#]As reported by *MolProbity* (Chen *et al.*, 2010)

6.3.3.4 Overall Topology of PA1672

PA1672 belongs to the class III βαβββ-module proteins of *P. aeruginosa* and to the *Pfam* clan of glyoxalases and shows the typical overall topology of this protein family (Figure 31). PA1672 folds into two βαβββ-motifs to form the full monomer. Both modules display the typical βαβββ structure and are numbered from β1-β8 and α1-α2 (Figure 31A). The protein forms a dimer in solution and in the crystal's asymmetric unit (Figure 31B). Dimer formation occurs by strong hydrophobic and salt-bridge interactions and is essential to form the potential ligand or substrate binding pockets at the two dimer interfaces of the PA1672 dimer. The PA1672 dimer is formed by edge-to-edge dimerization interactions, where β-sheets of the two monomers interact with each other, but α-helices of the same protein interact (Figure 31C).

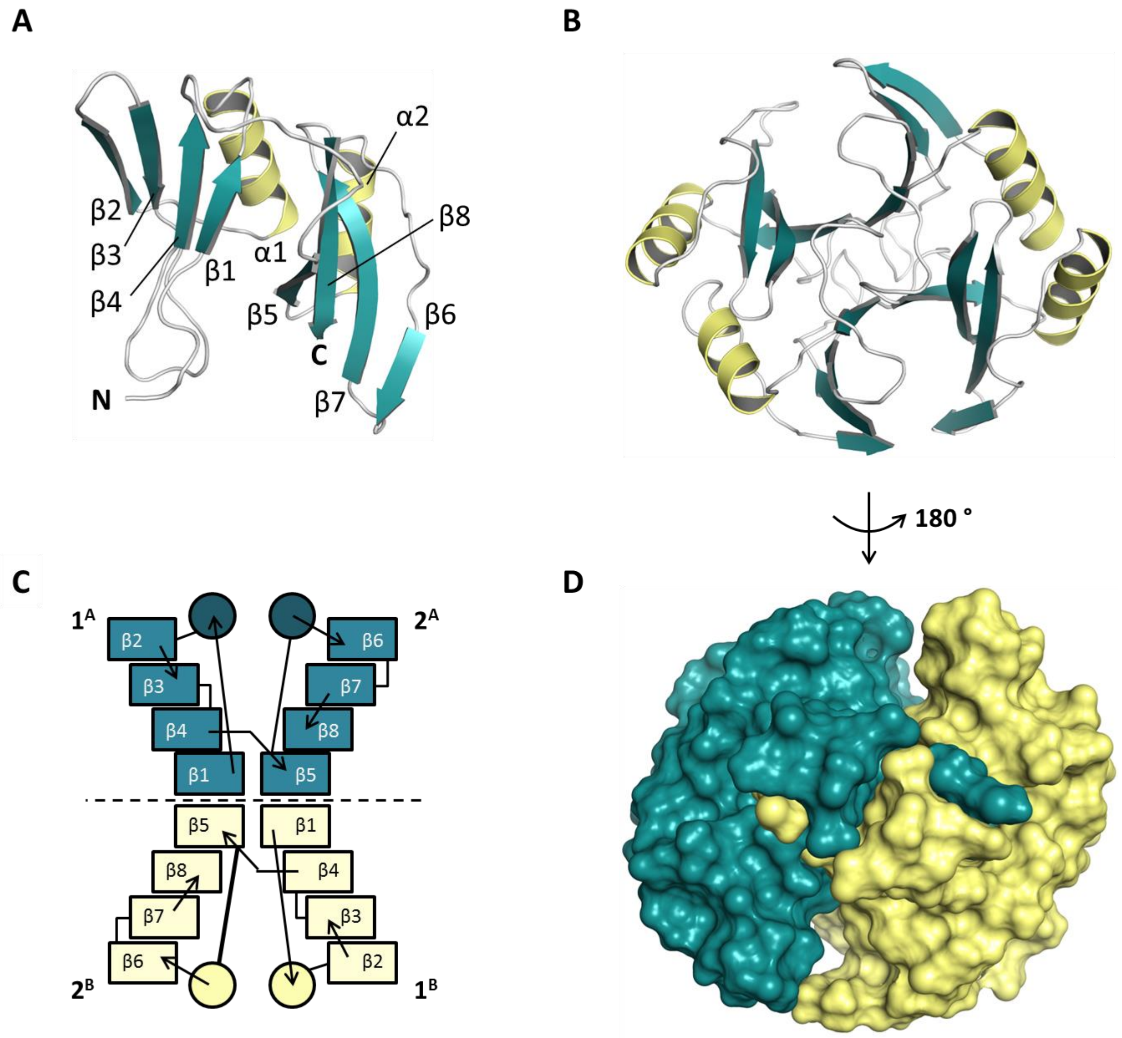


Figure 31: Overall topology of PA1672. A: The PA1672 monomer displays the typical βαβββ-module protein structure with two modules per chain. B: PA1672 forms a physiological dimer with two binding pockets. These are formed by both proteins of the homodimer. The homodimer is also found in the asymmetric unit of the PA1672 crystal. C: Topology diagram of PA1672. The PA1672 dimer is formed by edge-to-edge contacts. D: Surface representation of the nearly perfectly globular PA1672. The dimer shows arm-to-arm exchange that is stabilized by closure of interacting loops.

A surface representation of the PA1672 dimer clearly demonstrates dimerization via arm-to-arm exchange, a typical feature of βαβββ-module proteins. The arm-to-arm exchange is stabilized by two interacting loops that close over the N-terminus (Figure 31D). However, domain swapping is no observed in PA1672. The binding pocket of PA1672 is rather open and large (Figure 31B+D) and may possibly harbor larger molecules than other βαβββ-module resistance proteins.

6.3.3.5 Potential Active Center of PA1672

In order to identify potential catalytic residues, the probable substrate binding sites of PA1672 are examined (Figure 32). In the PA1672 potential catalytic pocket, several potential catalytic residues are located. For instance, histidines such as H6 and H42 often play a catalytic role or chelate metal cations (Bartlett *et al.*, 2002). Furthermore, several polar residues are found in the binding pocket, such as E8, or E107, S38, R115 and Q119. These residues might either be involved in hydrogen bonds or in catalytic activity of PA1672. Aromatic residues such as Y62, Y72 or F94 may either aid in establishing π-interactions with aromatic or hydrophobic compounds, or, in case of the tyrosines, may be catalytically involved. Also, M40, I70, L99 or I105 are able to engage in hydrophobic contacts with potential ligands. However, only from the presence of these amino acids in the binding site, no function of the protein can be deduced.

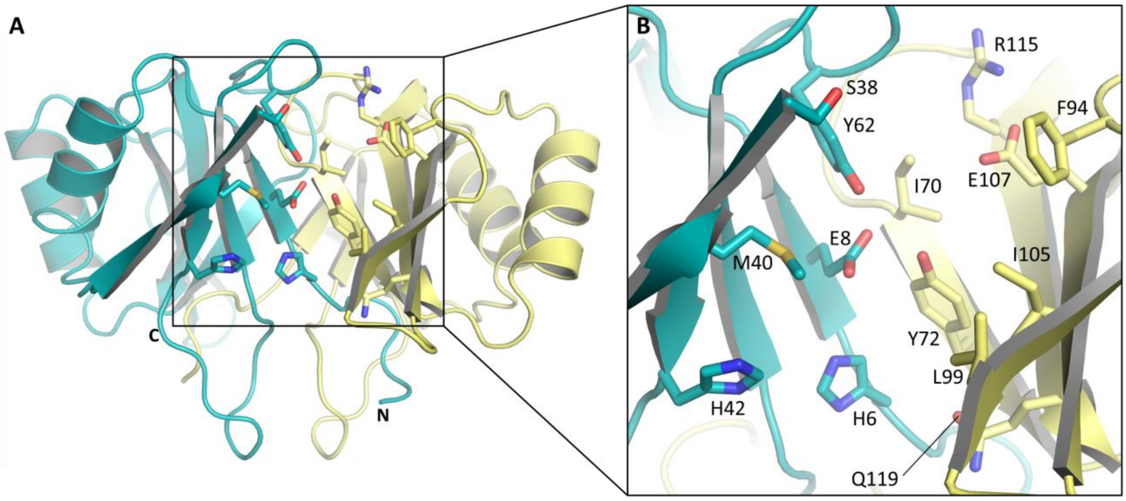


Figure 32: Potential active center of PA1672. A: Location of the ligand/substrate binding site in the PA1627 dimer. The binding site is formed by residues from both molecules of the dimer. B: Close-up of the potential active site with amino acid residues that might be involved in ligand/substrate binding or catalysis of the enzymatic reaction.

For this reason, the structure of PA1672 was submitted to the 3D Search for Protein SITEs (*SPRITE*) server (Nadzirin *et al.*, 2012) to analyze 3D motifs in the active site of PA1672 and identify a possible function for PA1672. The *SPRITE* server returned 17 distinct protein motif hits derived from PDB structures. The only two hits involving residues in the probable catalytic pocket of PA1672 are given in Table 26. No protein of the glyoxalase superfamily, but only proteins from unrelated fold families were identified by

SPRITE. For this reason, it is unlikely that PA1672 carries out the exact reaction that one of these proteins is able to catalyze. PA1672 might catalyze either a hydrolase or transferase reactions. In contrast, many of the already classified βαβββ-module proteins are oxidoreductases or lyases.

Table 26: Two hits of the 3D motif search with PA1672 using the *SPRITE* server (Nadzirin *et al.*, 2012). Only these hits involve residues in the potential binding pocket of PA1672.* = residues in the potential active center of PA1672.

PDB Hit	Description	r.m.s.d.	No. of residues	Matches Pattern | PA0880
1BGL	β-galactosidase	1.15	3	E537 | E8* Y503 | Y62* E461 | E107*
1ULA	purine nucleoside phosphorylase	1.29	3	H86 | H6* E89 | E8* N243 | N74*

6.3.4 Class III βαβββ-Module Protein PA4518

6.3.4.1 Bioinformatic Analysis of PA4518

A *BLAST* search (E-value cut off 0.001) in the PDB using the sequence of PA4518 identified two homologous proteins, a glyoxalase-related enzyme from *Fulvimarina pelagi* (PDB entry 3BQX, 37% sequence identity) and a glyoxalase/bleomycin resistance protein/dioxygenase from *Conexibacter woesei* (PDB entry 4GYM, 32% sequence identity). However, the proteins lack functional characterization and do not provide a starting point for functional characterization of PA4518.

In the *Phyre*^{2} structure prediction, 120 potential template structures were identified, and a high confidence PA4518 structure was modeled. PA4518 was correctly identified as a βαβββ-module protein. The top five hits of the *Phyre*^{2} modelling (Table 27) include the two hits identified by the *BLAST* search in the PDB, and also identify a mitomycin resistance protein from *S. lavendulae*, a metallo-protein from *Bacillus cereus* and the 2,3-dihydroxybiphenyl-1,2-dioxygenase (DHBD). Thus, PA4518 is homologous to metal-binding and aromatic compound binding proteins, indicating that no hint to a certain function is provided.

Table 27: Models used for *Phyre*^{2} predictions of PA4518. PF00903 = glyoxalase I, PF12681 = glyoxalase II. * = also identified in the *BLAST* search in the PDB.

PDB ID	Functional Annotation	Sequence Identity / %	Model Confidence / %	PFAM family
3BQX*	glyoxalase-related enzyme from *Fulvimarina pelagi*	31	100	PF00903, PF12681
4GYM*	glyoxalase/bleomycin resistance protein/dioxygenase from *Conexibacter woesei*	22	99.9	PF12681
1KLL	Mitomycin resistance protein from *S. lavendulae*	20	99.9	PF00903, PF12681
1ZSW	*Bacillus cereus* metallo-protein	18	99.9	PF00903
1LGT	2,3-dihydroxybiphenyl-1,2-dioxygenase (DHBD)	16	99.9	PF00903

Conserved domain analysis with *CDD* (Marchler-Bauer *et al.*, 2015) correctly identifies PA4518 as a member of the glyoxalase superfamily, specifically as a member of glyoxalase I/bleomycin resistance protein family. Furthermore, it predicts the protein to bind a divalent metal cation. However, in the sequence alignment with *PROMALS3D* (Figure 22 and Figure 23), only one of the three metal-binding residues seems to be conserved in PA4518. Similarly, the aromatic amino acid side chains used for π-stacking interactions with aromatic moieties are not conserved in PA4518. This indicates that PA4518 might have enzymatic function that does not require divalent metal cations, or that it does bind compounds that lack aromatic moieties.

6.3.4.2 Crystallization and Data Collection of PA4518

Nineteen initial hits were identified for PA4518, but most of them were tiny needles of 30 µm length or less. Three of the initial hits (JCSG Core II C8, Core III F6 and Core IV F9) were optimized. Finally, well-diffracting crystals were obtained from 0.1 M sodium citrate pH 3.6, 10% (*w/v*) MPD (Figure 33).

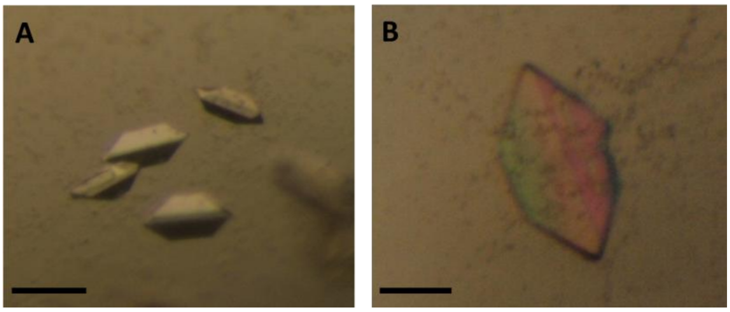


Figure 33: Crystals of PA4518 in 0.1 M sodium citrate pH 3.6, 10% (w/v) MPD. A: Crystals from the screen B: Optimized crystals. Scale bar is 200 µm.

Crystals of PA4518 were cryoprotected in mother liquor supplemented with additional 15% (v/v) MPD before flash cooling them in liquid nitrogen. Diffraction data a PA4518 crystal was collected as 100 1° images on a MarCCD Chess detector at a wavelength of λ = 0.9184 Å on beamline 14.1 at BESSY II Synchrotron of the HZB, Berlin, Germany. Data collection and processing statistics are given in Table 28.

Table 28: Data collection and processing statistics for *P. aeruginosa* PAO1 PA4518. Values in parentheses are for the highest resolution shell.

Dataset	PA4518
Wavelength (Å) / beamline≠	0.9184 / BESSY II, 14.1
Resolution range (Å)	46.08-1.47 (1.5-1.47)
Space group	$P2_12_12_1$
Unit cell parameters (Å) (°)	42.0 72.73 92.16 90 90 90
Mosaicity (°)†	0.143
Total No. of measured reflections	198136 (9644)
Unique reflections	48687 (2362)
Multiplicity	4.1 (4.1)
Mean I/σ(I)	11.9 (2.0)
Completeness (%)	99.7 (99.4)
R_{meas} (%)	9.7 (85.8)
$R_{p.i.m.}$ (%)	4.7 (41.5)

†Mosaicity values reported by *XDS* (Kabsch, 2010)

The PA4518 crystal diffracted to a resolution of 1.47 Å at the BESSY II Synchrotron and was indexed in orthorhombic space group $P2_12_12_1$ with cell parameters a = 42.0, b =

72.73 and c = 92.16. The Matthews coefficient (Matthews, 1968) was calculated as 2.29 Å^3/Da which corresponds to a solvent content of 46% and two PA4518 molecules (one dimer) in the asymmetric unit.

6.3.4.3 Structure Determination and Refinement of PA4518

The structure of PA4518 was determined by MR in *PHASER* from *CCP4* using a dimer of the crystal structure of an uncharacterized glyoxalase-related enzyme from *Fulvimarina pelagi* (PDB entry 3BQX, Rao *et al.*, to be published) with 37% sequence identity.

Table 29: Refinement statistics for PA4518. Values in parentheses are for the highest resolution shell.

Dataset	PA4518
Resolution range (Å)	46.08-1.47 (1.5-1.47)
R_{work} (%)	13.6 (22.5)
R_{free} (%)	15.6 (23.5)
No. of non-H atoms	
Protein	2630
Ion	-
Ligand	32
Water	411
R.m.s. deviations	
Bonds (Å)	0.012
Angles (°)	1.453
Average B factors (Å^2)	
Protein	14
Ion	-
Ligand	36
Water	30
Ramachandran plot	
Favored regions (%)	98
Outliers (%)	0
MolProbity score[#]	1.27

[#] As reported by *MolProbity* (Chen *et al.*, 2010)

The structure of PA4518 was obtained with an R_{work} = 13.6% and an R_{free} = 15.6% with 98% of the residues in Ramachandran-favored regions and a *MolProbity* score of 1.27 (Table 29).

6.3.4.4 Overall Topologies of PA4518

PA4518 belongs to the class III βαβββ-module proteins of *P. aeruginosa* and to the *Pfam* clan of glyoxalases and folds into two typical βαβββ-motifs (Figure 34).

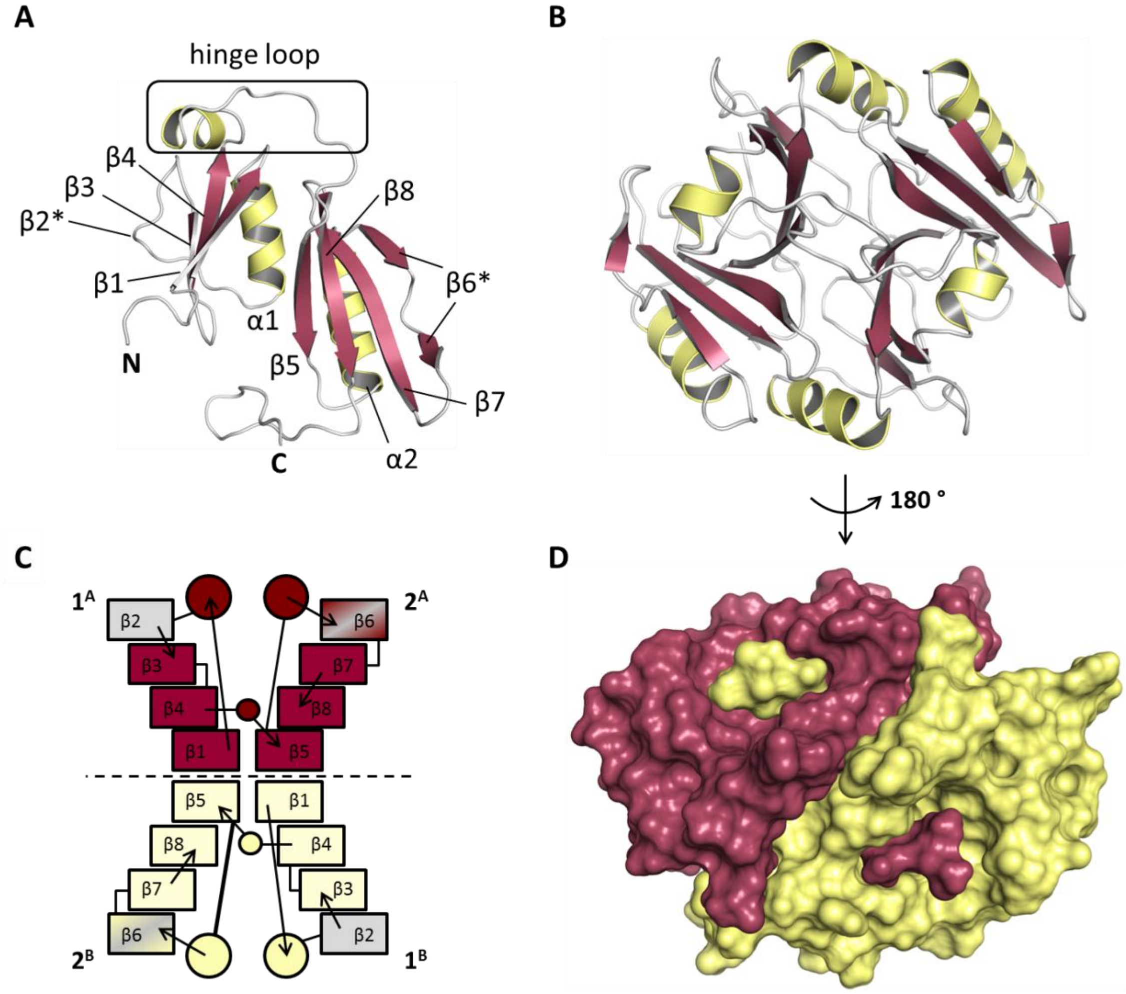


Figure 34: Overall topology of PA4518. A: The PA4518 monomer consists of two βαβββ-modules. However, β2 and β6 are degenerate and rather unstructured. Interestingly the protein carries a hinge loop with an additional α-helical element. B: PA4518 forms a dimer in solution as well as in the asymmetric unit of the crystal. C: The topology diagram of PA4518 displays the dimerization via edge-to-edge contacts. D: Surface representation of PA4518 shows tight arm-to-arm exchange between the two monomers. The dimerization is further stabilized by interacting loops that close across the N-terminus.

However, β2 and β6 seem to be rather unstructured or only partially structured compared to the β-sheets in PA1672. Interestingly, the protein carries a hinge loop with an additional α-helix. PA4518 forms a highly stable dimer in solution and in the crystal's asymmetric unit (Figure 34B). The PA4518 dimer is formed by edge-to-edge interactions, where β-sheets of the two monomers interact with each other, but α-helices of the same monomer interact (Figure 34C). At the dimer interfaces, the potential ligand or substrate binding pockets is formed. This pocket is smaller than the pocket of PA1672. A surface representation of the PAPA4518 dimer (Figure 34D) clearly demonstrates dimerization via arm-to-arm exchange, a typical feature of βαβββ-module proteins. The arm-to-arm exchange is stabilized by two loops from the other protein which interact and close over the N-terminus.

6.3.4.5 Potential Active Center of PA4518

PA4518 displays two different active site conformations, an open form and a closed form (Figure 35A). In the open form, the loop Q103 to Y110 sticks out of the protein with W107 pointing towards a neighboring molecule in the crystal. This conformation allows the binding pocket to harbor two MPD molecules from the mother liquor (Figure 35D). In the closed conformation, this loop Q103 to Y110 turns towards the binding pocket of PA4518. In this conformation, W107 is turned into the binding pocket, which causes it to be smaller and to contain only one MPD molecule (Figure 35C). The conformational differences are caused by crystal packing, but might as well be of physiological relevance.

As predicted by the sequence alignment with *PROMALS3D* (Figure 22 and Figure 23), no metal-binding site is observed in this protein. Also, no evident aromatic compound binding motif is found in PA4518. Still, a variety of residues potentially involved in ligand binding or catalytic activity is located in the PA4518 binding pocket. For instance, aromatic residues such as F40, Y51, Y113, W 107, F129 and W130 may either aid in establishing π-interactions with aromatic or hydrophobic compounds, or, in case of the tyrosines, may be catalytically involved. Especially W107 might play a role in ligand/substrate binding, as it is located in the flexible loop of PA4518. Similarly, L8, V38, L56 and V105 are able to engage in hydrophobic contacts. Polar residues such as T10, D59, T73, N103, S111, E123 and N127 might either establish hydrogen bonds or in catalytic activity of

PA4518. However, no cysteine or histidine residues which are often involved in catalytic activity (Bartlett *et al.*, 2002) are found in the binding pocket of PA4518.

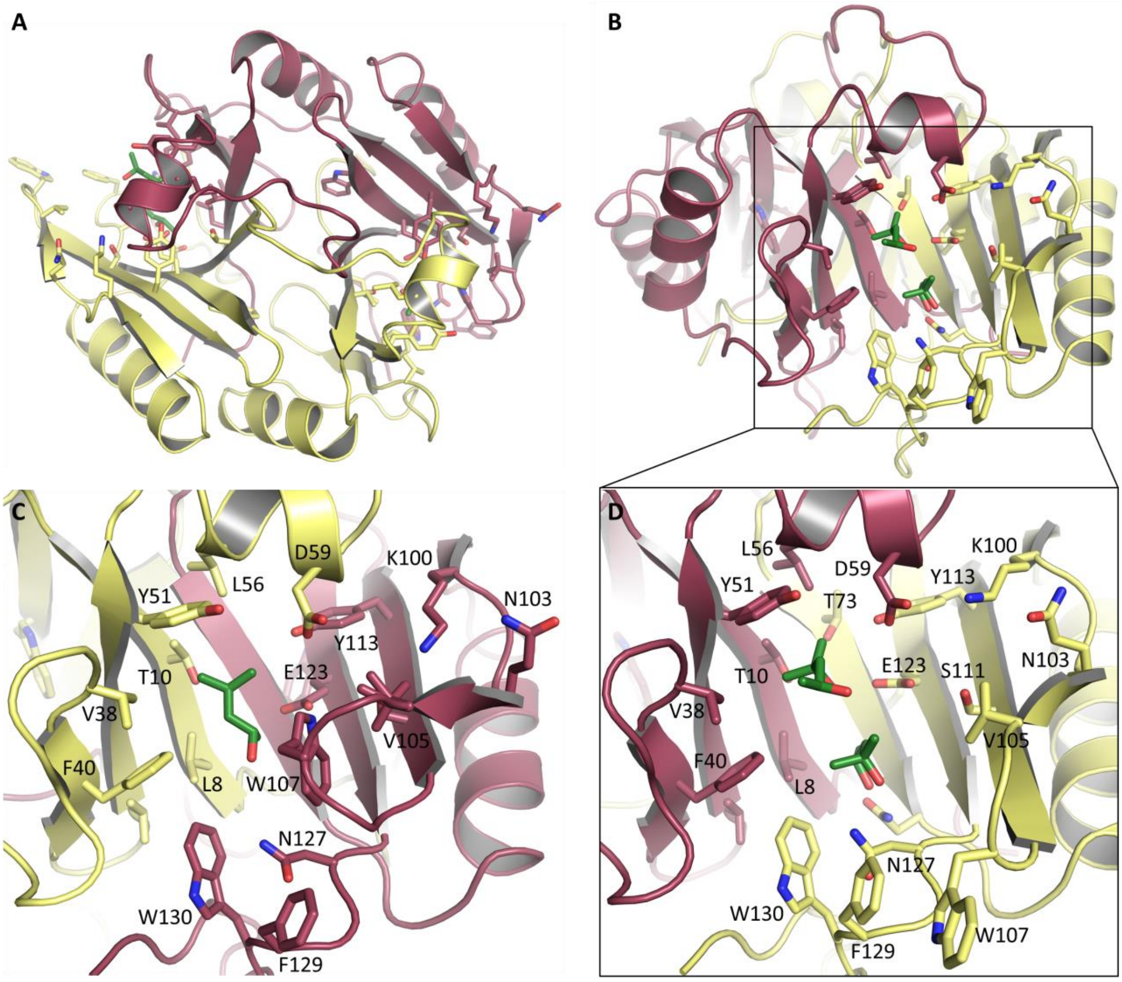


Figure 35: Potential active center of PA4518 with MPD (green) from the precipitant solution. A: Top view of PA4518 with an open (left) and closed (right) conformation of the binding pocket. B: Location of the ligand/substrate binding site in the PA4518 dimer. C: Detailed view of the binding pocket in its closed conformation with amino acid residues that might be involved in ligand/substrate binding or catalysis of the enzymatic reaction. D: Detailed view of the binding pocket in its open conformation.

The structure of PA4518 was submitted to the *SPRITE* server (Nadzirin *et al.*, 2012) to analyze 3D motifs in the active site of the protein and identify a possible function for PA4518. Analysis with the *SPRITE* server returned 31 distinct protein motif hits derived from PDB structures. However, none of these hits involves side chains in the binding pocket of PA4518. Similarly, analysis of PA4518 with *PDBeFold* (Krissinel & Henrick, 2004) identified 180 hits, but as the fold of this families is highly conserved, while the sequence is not, this analysis did not aid in identifying a possible function for PA4518.

6.4 The Probable Itaconate-Metabolic Enzyme PA0880

6.4.1 Bioinformatic Analysis of PA0880

In this study, PA0880 was produced as a His_6-PA0880 fusion protein with a calculated molecular weight of 15.75 kDa (Gasteiger *et al.*, 2003). After TEV cleavage, two residues GH are left on the N-Terminus. The resulting protein is a 128-amino acid protein with a calculated molecular weight of 13.87 kDa, a low extinction coefficient of 2980 $M^{-1}cm^{-1}$ and a theoretical pI of 5.28 (Gasteiger *et al.*, 2003). The amino acid sequence of PA0880 was submitted the PDB, which returned three similar structures, all of them βαβββ-module proteins. Additionally, the sequence of PA0880 was sent to the structure prediction server *Phyre*^{2}. *Phyre*^{2} returned a structural model with high confidence, but no hints to a certain protein function. The top five hits, including the three structures identified by the *BLAST* search in the PDB are given in Table 30.

Table 30: Models used for *Phyre*^{2} predictions of PA0880. PF12681 = glyoxalase II family, PF00903 = Glyoxalase/Bleomycin resistance protein family.*= also identified in PDB search

PDB ID	Functional Annotation	Sequence Identity / %	Model Confidence / %	PFAM family
3ZW5*	human glyoxalase domain protein 5	57	99.9	PF12681
3HUH*	biphenyl-2,3-diol 1,2-dioxygenase iii-related protein2 from *S. typhimurium*	65	99.9	PF12681
3EY7*	mobile metagenome of *V. cholera* integron cassette protein vch_cass1	60	99.9	PF12681
1ZSW	*Bacillus cereus* metallo protein from glyoxalase II family	21	99.9	PF00903
4HC5	glyoxalase/bleomycin resistance protein from *Sphaerobacter thermophilus*	17	99.9	PF00903, PF12681

Homologs of the *P. aeruginosa* class III βαβββ-module protein PA0880 belong to the glyoxalase I/bleomycin resistance protein family (PF00903) and to the glyoxalase II protein family (PF12681). However, no hint pointing to a certain function can be derived from the structural homolog predictions or the protein families. Also, no publications on these five proteins are available to retrieve hints to the protein functions. Furthermore, *BLAST* searches, domain analysis or secondary structure predictions all yield more than 8000 hits, making it impossible to identify a protein with the same or very similar function to PA0880 just from sequence homology. Interestingly, the cytoplasmic βαβββ-module protein PA0880 is predicted to be encoded in an operon with PA0878, PA0879, PA0881, PA0882 and PA0883 (Mao *et al.*, 2009). This gene cluster has recently been demonstrated to be involved in itaconate metabolism in *P. aeruginosa* (Sasikaran *et al.*, 2014).

6.4.2 Background - Itaconate Production and Metabolism

Itaconate (methylene succinate) is a mammalian metabolite produced in phagocytic cells such as macrophages. These cells enclose the invading bacterial cell to form the phagosome, and later, the phagolysosome, where the bacterium is degraded. Macrophages are considered as part of the primary line of defense upon a bacterial infection. Recently, macrophages were shown to produce and secrete itaconate in response to increased interferon γ levels and lipopolysaccharide (Strelko *et al.*, 2011). Itaconate production is therefore considered as an antibacterial response of these phagocytic cells (Strelko *et al.*, 2011). This is because itaconate is a potent inhibitor of isocitrate lyase (McFadden & Purohit, 1977), a key enzyme in the glyoxalate cycle. Many pathogenic bacteria use the glyoxalate cycle to assimilate AcCoA produced during fatty acid degradation (Lorenz & Fink, 2002) and fatty acids seem to be the main carbon source in macrophages. Additionally, itaconate is secreted and used as a carbon source by extracellular pathogens such as *P. aeruginosa* (Mittal *et al.*, 2009). Thus, the ability of pathogenic bacteria to metabolize itaconate is regarded as an evolutionary advantage that helps the bacteria to survive and persist in the host cell.

The itaconate degradation pathway of *P. aeruginosa* is slightly different from the one of *Yersinia pestis*. Both pathways have recently been discovered and investigated in the same study (Sasikaran *et al.*, 2014). In the first step of itaconate degradation in *P. aeruginosa*, itaconate is activated to the corresponding CoA ester by succinyl-CoA:itaconate CoA transferase (Ict) (Figure 36). Then, itaconyl-CoA is hydrated to (*S*)-citramalyl-CoA by an itaconyl-CoA hydratase (Ich). Finally, (*S*)-citramalyl-CoA is cleaved to AcCoA and pyruvate by an (*S*)-citramalyl-CoA lyase (Ccl). In *P. aeruginosa*, the genes encoding the itaconate-degrading enzymes were recently identified via a *BLAST* search with the *Y. pestis* Ccl (Sasikaran *et al.*, 2014). The *P. aeruginosa* Ccl (PA0883) is encoded in an operon with

Ict (PA0882) and Ich (PA0878). Surprisingly, three additional proteins are encoded in this operon: a putative acyl-CoA dehydrogenase (PA0879), an MmgE-PrpD family protein (PA0881) and the βαβββ-module protein (PA0880) (Sasikaran *et al.*, 2014).

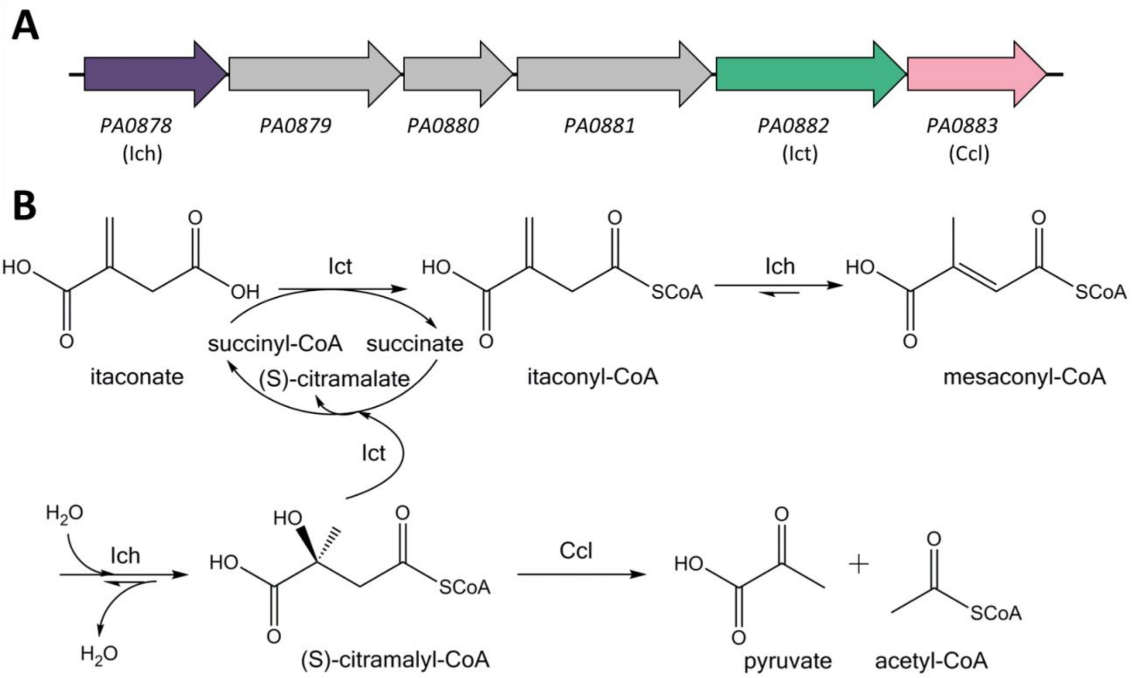


Figure 36: Scheme of the itaconate degradation pathway in *P. aeruginosa* (Sasikaran *et al.*, 2014). Itaconate is activated to itaconyl-CoA by Ict under succinyl-CoA consumption. Ich transforms itaconyl-CoA into mesaconyl-CoA and hydrates it to form (*S)*-citramalyl-CoA. Finally, Ccl degrades (*S*)-citramalyl-CoA into AcCoA and pyruvate.

Since *Pseudomonas* spp. are also able to metabolize (*S*)-citramalate, mesaconate and methylsuccinate, it is tempting to speculate that the three uncharacterized enzymes PA0879, PA0880 and PA0881 might be responsible for channeling these substrates into the itaconate degradation pathway (Sasikaran *et al.*, 2014). After potential formation of methylsuccinyl-CoA by Ict, PA0879 might catalyze the methylsuccinyl-CoA dehydrogenase reaction, while PA0881 might convert mesaconate into (S)-citramalate, because mesaconate cannot be activated to the CoA ester by Ict (Sasikaran *et al.*, 2014). However, a potential function of PA0880 remains elusive. PA0880 might be involved in conversion of the product of the PA0879 reaction, or conversion of (*S*)-citramalate to a compound that can be activated by Ict. Assuming that PA0880 might catalyze a standard glyoxalase reaction, the protein might act as an isomerase or epoxide-opening enzyme (Figure 37, Armstrong, 2000). Hence, it could catalyze an isomerization reaction on (*S*)-

citramalate, or an epimerization reaction on a CoA derivative, such as the methylmalonyl-CoA epimerase. Still, such enzymes require methyl ions.

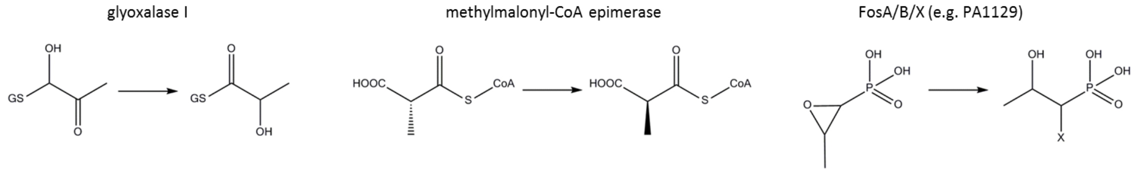


Figure 37: Isomerization and epoxide-opening reactions catalyzed by the metal-depending clade of the βαβββ-module proteins. However, PA0880 is predicted not to contain a metal-ion. GS = glutathione

Another possibility could be a nucleophilic attack of a cysteine residue at the carbonyl-CoA of a CoA derivative. βαβββ-module proteins such as the fosfomycin resistance proteins FosA/B/X (e. g. PA1129) catalyze such nucleophilic attacks in the opening of the epoxide ring of fosfomycin.

6.4.3 Crystallization and Data Collection of PA0880

PA0880 was screened with the JCSG Core Screens I-IV (Qiagen). Several conditions yielded needle-shaped crystals. A well-diffracting crystal was obtained directly from the screening plate in 0.1 M sodium acetate pH 4.5, 50% (*v/v*) ethylene glycol and 5% (*w/v*) PEG 1000 (JCSG Core II F3) (Figure 38). However, it was not possible to reproduce and optimize crystals of PA0880 in any condition tested on 24-well plates.

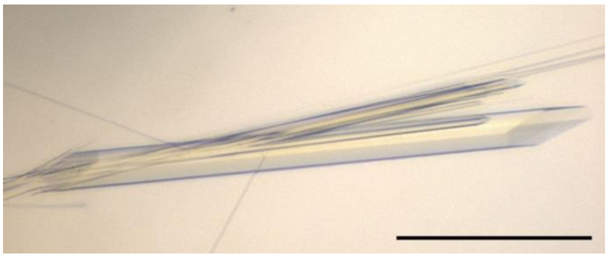


Figure 38: Crystals of PA0880 from JCSG Core II F3. The crystals could not be reproduced in an optimization plate and therefore not be optimized. Scale bar is 200 µm.

Diffraction data collection of the PA0880 crystal required no further cryoprotection and was performed on beamline 14.1 at BESSY II Synchrotron of the HZB, Berlin, Germany, where 100 images of 1° oscillation were collected on a MarCCD Chess detector at a wavelength of λ = 0.9184 Å. Data collection statistics are given in Table 31.

Table 31: Data collection and processing statistics of *P. aeruginosa* PAO1 PA0880. Values in parentheses are for the highest resolution shell.

Dataset	PA0880
Wavelength (Å) / beamline[≠]	0.9184 / BESSY II, 14.1
Resolution range (Å)	45.54-1.53 (1.56-1.53)
Space group	$P2_12_12_1$
Unit cell parameters (Å)	42.86 49.93 111.03
(°)	90 90 90
Mosaicity (°)[†]	0.091
Total No. of measured reflections	148584 (7214)
Unique reflections	36757 (1772)
Multiplicity	4.0 (4.1)
Mean I/σ(I)	15.1 (3.0)
Completeness (%)	99.9 (99.6)
R_{meas} (%)	7.2 (56.2)
$R_{p.i.m.}$ (%)	3.6 (27.5)

[†] Mosaicity values reported by *XDS* (Kabsch, 2010)

The PA0880 crystal diffracted to a resolution of 1.53 Å and was indexed in orthorhombic space group $P2_12_12_1$ with cell parameters a = 42.86, b = 49.93 and c = 111.03. The Matthews coefficient (Matthews, 1968) was calculated as 2.16 Å^3/Da which corresponds to a solvent content of 43% and two PA880 molecules (one dimer) in the asymmetric unit.

6.4.4 Molecular Replacement and Refinement of PA0880

An appropriate MR model for the structure of PA0880 was identified by a *BLAST* search with the PA0880 sequence in the PDB. The structure of PA0880 was then solved using *PHASER* (McCoy *et al.*, 2007) using a dimer of the virulence protein STM3117 from *Salmonella typhimurium* (PDB entry 3HNQ, Northeast Structural Genomics Consortium, target ID StR274, Seetharaman *et al.* to be published). This protein shares 66% sequence identity with PA0880, and has also been identified in the previous *BLAST* and *Phyre*^{2} search in the PDB. It is equivalent to a biphenyl-2,3-diol 1,2-dioxygenase iii-related protein2 from *S. typhimurium* (PDB entry 3HUH, Northeast Structural Genomics Consortium, Fedorov *et al.*, to be published). The structure of PA0880 was finally refined to an R_{work} = 15.4% and an R_{free} = 17.6% with a *MolProbity* score of 0.92. In the PA0880 structure, 99.2 % of the residues are found in Ramachandran-favored regions (Table 32).

Table 32: Refinement statistics for *P. aeruginosa* PA0880. Values in parentheses are for the highest resolution shell.

Dataset	PA0880
Resolution range (Å)	45.54-1.53 (1.56-1.53)
R_{work} (%)	15.4 (20.5)
R_{free} (%)	17.6 (20.8)
No. of non-H atoms	
Protein	2256
Ion	12
Ligand	70
Water	308
R.m.s. deviations	
Bonds (Å)	0.004
Angles (°)	1.013
Average B factors (Å^2)	
Protein	14
Ion	25
Ligand	34
Water	34
Ramachandran plot	
Favored regions (%)	99.2
Outliers (%)	0
MolProbity score[#]	0.92
PDB entry code	-

[#] As reported by *MolProbity* (Chen *et al.*, 2010)

6.4.5 Overall Topology of PA0880

According to homology predictions, PA0880 belongs to the family of glyoxalase II or βαβββ-module resistance proteins (Kanehisa, 2002). In the crystal, the PA0880 monomer folds into two canonical βαβββ-motifs with the first motif reaching from amino acid residues 7 to 52 and the second motif from 69 to 125 (Figure 39A). The protein forms a dimer in solution and in the crystal’s asymmetric unit (Figure 39B). The PA0880 dimer is formed by edge-to-edge interactions, where β-sheets β1 and β5 of the two monomers interact with each other, but α-helices interact intramolecularly (Figure 39C). A surface representation of the PA0880 dimer clearly demonstrates dimerization via arm-to-arm exchange, a typical feature of βαβββ-module proteins.

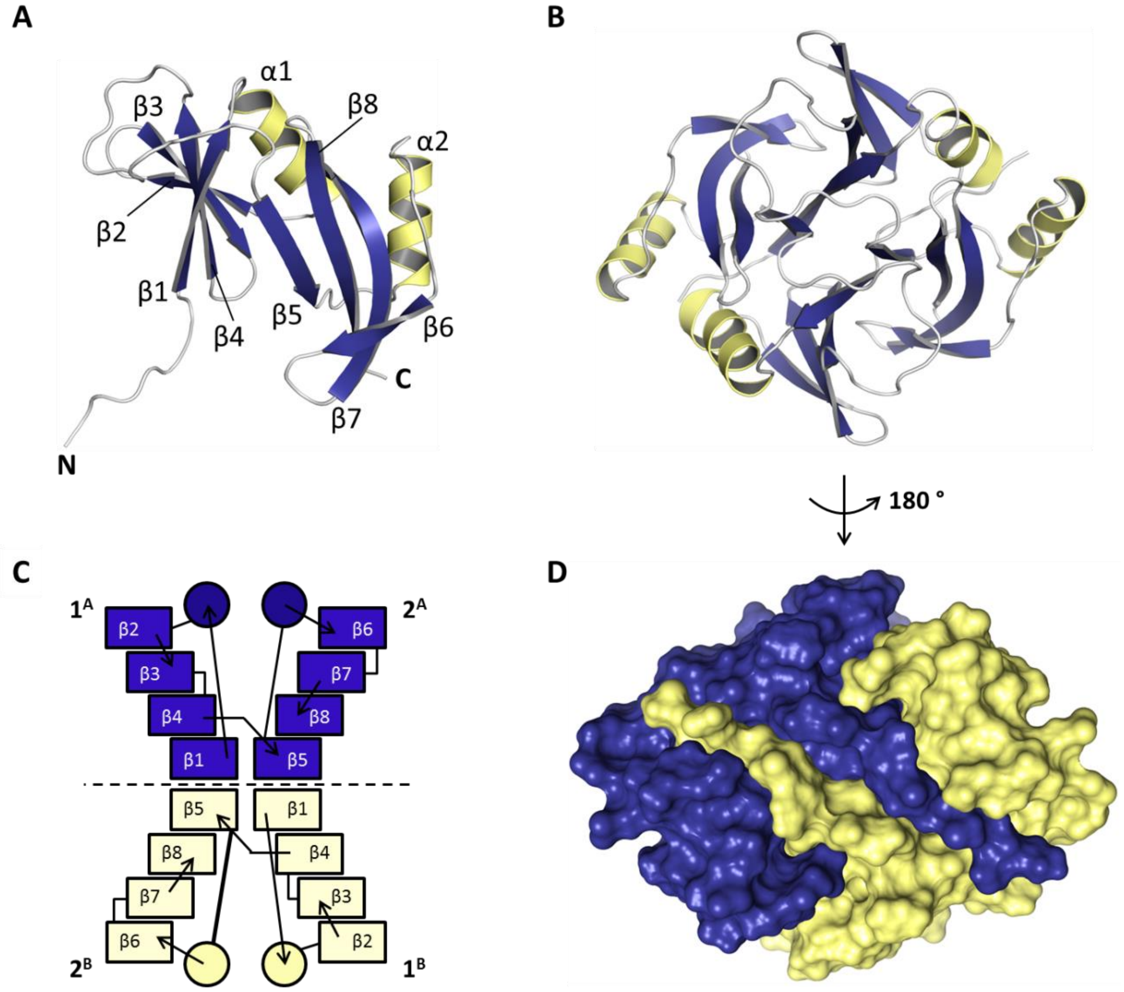


Figure 39: Overall topology of PA0880. A: PA0880 monomer shows the characteristic architecture with two βαβββ-modules. B: Physiological dimer of PA0880 as it is found in the asymmetric unit of the crystal. C: Topology diagram of the PA0880 dimer which is formed by edge-to-edge contacts. D: Surface representation of the PA0880 dimer shows significant arm-to-arm exchange of the two PA0880 proteins forming the homodimer.

The N-terminus of the first protein chain stretches across the whole second protein chain (Figure 39D) and vice versa, thereby providing strong dimer interactions. These interactions are typically found in proteins of the βαβββ-module protein family.

6.4.6 ITC Experiments with PA0880

As PA0880 was suspected to be involved in conversions of methylsuccinate or (*S*)-citramalate in the itaconate degradation pathway (Sasikaran *et al.*, 2014), twenty potential substrates (Figure 40) of PA0880 were selected based on the respective publication and examined for binding in ITC experiments.

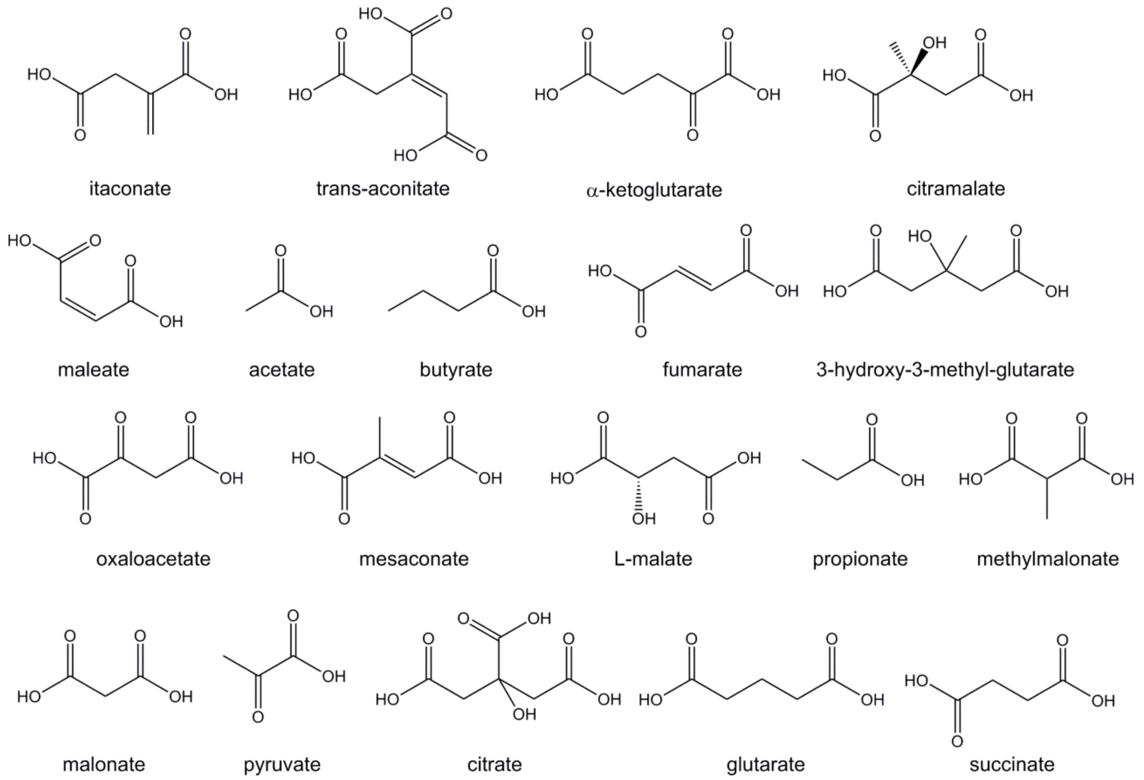


Figure 40: Potential ligands or substrates of PA0880 tested in ITC experiments. The protein showed no affinity to any of these potential ligands which were selected on the basis of a publication by Berg and coworkers (Sasikaran *et al.*, 2014). Pyocyanin was used as a control.

In the ITC experiments of PA0880, no affinity was detected for the potential twenty ligands or substrates. This either demonstrates that none of the substances tested is an actual ligand or substrate to PA0880, or that the conditions chosen for the experiment were not suitable to test these compounds. As the same conditions were also chosen for successful ITC experiments in this study, it is more likely that the tested substances do indeed not bind to PA0880. Another reason for the unsuccessful ITC experiments could be that PA0880 binds a CoA derivative of one of the ligands. These derivatives have not been tested yet in the ITC experiments. For this reason, no protein function can be derived from these experiments.

6.4.7 Potential Active Center of PA0880

The potential active center of PA0880 is located at the dimer interface and formed by residues from both molecules of the PA0880 dimer (Figure 41). In this binding pocket, two histidines H8 and H51 are located, as well as a cysteine C72. These residues are often involved in metal-binding or catalysis. However, no metal ion could be found in the binding pocket of PA0880. Further residues present are aromatic side chains as F57 or Y111, aliphatic amino acids as L119, and polar residues such as K47, N49, D70, R100, S109, E121 and S123.

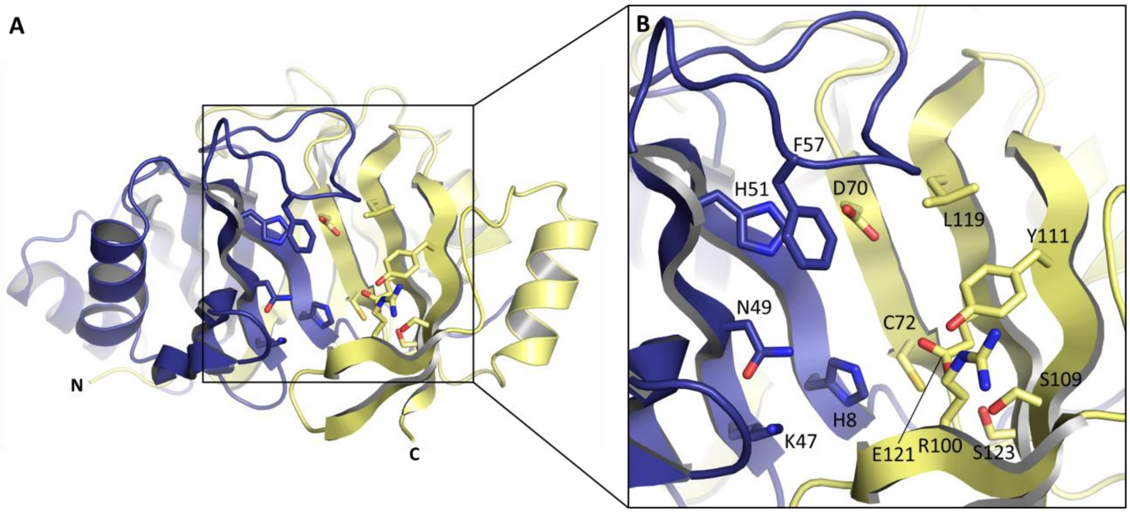


Figure 41: Potential active center of PA0880. A: Location of the ligand/substrate binding site in the PA0880 dimer. The binding site is formed by residues from both molecules of the dimer. B: Close-up of the potential active site with amino acid residues that might be involved in ligand/substrate binding or catalysis of the enzymatic reaction.

Histidine and cysteine are the amino acid residues with the highest catalytic propensity (Bartlett *et al.*, 2002). This means that they are much more abundant in active centers of enzymes than they are over the whole protein. For instance, the motif H8-C72-E121 in PA0880 could be a potential Zn(II)-binding motif (Auld, 2001). However, no metal ion is coordinated in the structure of PA0880. Cysteine is also often involved in nucleophilic attacks, as in hydrolases or proteases. One example for a cysteine protease is papain that utilized a histidine-activated cysteine as its catalytic residue.

The structure of PA0880 was submitted to the *SPRITE* server (Nadzirin *et al.*, 2012) to analyze 3D amino acid motifs in the active site of PA0880 and identify a possible function for PA0880. The *SPRITE* server returned 26 distinct protein motif hits derived from PDB structures. The top five hits are given in Table 33.

Table 33: Top five hits of the 3D motif search with PA0880 using the *SPRITE* server (Nadzirin *et al.*, 2012). Citrate synthase and propionyl-CoA carboxylase bind CoA derivatives.* = residues in the potential active center of PA0880.

PDB Hit	Description	r.m.s.d.	No. of residues	Matches Pattern | PA0880
1AJ8	citrate synthase	0.98	3	H262 | H8*, H223 | H51*, D312 | D70*
1XNY	propionyl-CoA carboxylase complex B	1.03	5	H262 | H8*, N465 | N49*, H223 | H51*, D312 | D70*, C118 | C72*
1VJV	ubiquitin carboxyl-terminal hydrolase	1.07	3	A420 | A45, G182 | G102, G183 | G104
1AL6	citrate synthase	1.07	3	H447 | H8*, N465 | N49*, C118 | C72*
1CT9	asparagine synthetase B	1.08	10	H320 | H8*, N175 | N49*, H274 | H51*, A420 | A53, G419 | G54, G183 | G55, G75 | G67, D375 | D70*, C25 | C72*, N74 | N118

However, no protein of the glyoxalase superfamily, but only proteins from unrelated fold families were identified by *SPRITE*. Of these top five hits, the ubiquitin carboxyl-terminal hydrolase and the asparagine synthetase B do not seem to be similar to PA0880, but might catalyze a similar type of reaction. Still, their function may also be completely unrelated to the one of PA0880, as PA0880 is encoded in the itaconate metabolism gene cluster of *P. aeruginosa*. In contrast, citrate synthase catalyzes the condensation of oxaloacetate and AcCoA, and propionyl-CoA carboxylase degrades propionyl-CoA to methylmalonyl-CoA and then succinyl-CoA. Oxaloacetate, however, has been tested in the ITC experiments, but was found to not bind to PA0880. Hence, CoA derivatives should be in the focus of future binding experiments. These experiments can be accelerated by testing metabolite mixtures of five compounds. Another possibility to probe the active site of PA0880 would be by docking of metabolite libraries.

6.5 Pyocyanin Resistance in *P. aeruginosa* PAO1

In previous studies, the crystal structures of the PA0803, PA1353 and PA46141 were determined using X-ray crystallography (Yu, 2009; Kalawy-Fansa, 2010). All three proteins were shown to bind pyocyanin in ITC experiments and were therefore assigned to be potential pyocyanin resistance factors. Crystals of PA0803, PA1353 and PA4641 were soaked with pyocyanin, and ligand complexes of PA1353 and PA4641 with pyocyanin were obtained (Yu, 2009; Kalawy-Fansa, 2010). However, pyocanin was only found in one of the four binding sites in the asymmetric unit of PA0803 crystals and the quality of the electron density indicated that the ligand was only loosely bound. For this reason, cocrystallization of PA0803 with pyocyanin was pursued during this study in order to obtain a crystal structure of PA0803 with 100% pyocyanin occupancy.

6.5.1 Cocrystallization and Data Collection of PA0803 with Pyocyanin

PA0803 was preincubated with 1 mM pyocyanin and then screened with PA0803 alone as a control. Crystals appeared in 14 different conditions. However, twelve of them contained high concentrations of small PEGs (PEG 200, PEG 300, PEG 400, PEG 600) as the apo condition which has been used for the soaking approach (Yu, 2009). For this reason, these conditions were discarded, as a PEG 200 molecule occupies the probable pyocyanin binding site in the apo structure (Yu, 2009). Only two conditions did not contain small PEGs, however, only one of them yielded crystals only in the presence of pyocyanin, but no crystals in the absence of pyocyanin (Core IV H6). These crystals were optimized in a 24-well setup. Finally, rod-shaped crystals were obtained from 0.8 M sodium dihydrogen phosphate, 1.2 M dipotassium hydrogen phosphate, 0.1 M sodium acetate pH 4.5, 2.5 mM pyocyanin and 2.5 mM DMSO (Figure 42). Crystals were cryoprotected in mother liquor supplemented with 15% (*w/v*) glycerol and flash cooled in liquid nitrogen.

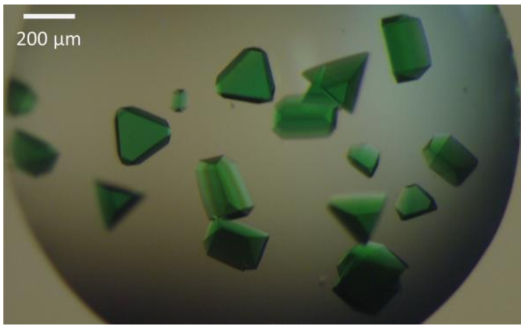


Figure 42: Rod-shaped green crystal of PA0803 cocrystallized with pyocyanin.

Well-diffracting PA0803 crystals with pyocyanin were sent to the BESSY II Synchrotron of the HZB, Berlin. Diffraction data of a PA0803 crystal was collected as 200 images of 0.5° rotation at a wavelength of λ = 0.9184 Å (Table 34).

Table 34: Data collection and processing statistics for *P. aeruginosa* PA0803 in complex with pyocyanin. Values in parentheses are for the highest resolution shell.

Dataset	PA0803 + PYO
Wavelength (Å) / beamline	0.9184 / BESSY, BL14.1
Resolution range (Å)	46.39-1.71 (1.74-1.71)
Space group	$P2_12_12$
Unit cell parameters (Å) (°)	84.26 101.8 92.77 90 90 90
Mosaicity (°)†	0.109
Total No. of measured reflections	357368 (19052)
Unique reflections	86649 (4588)
Multiplicity	4.1 (4.2)
Mean I/σ(I)	11.6 (2.1)
Completeness (%)	99.9 (99.9)
R_{meas} (%)	9.1 (81.2)
$R_{p.i.m.}$ (%)	4.5 (39.3)

†Mosaicity values reported by *XDS* (Kabsch, 2010)

The PA0803-pyocyanin crystal diffracted to 1.71 Å and was indexed in orthorhombic space group $P2_12_12$ with cell parameters a = 84.26, b = 101.8 and c = 92.77. Subsequently, these data were used to determine the PA0803 pyocyanin complex.

6.5.2 Structure Determination and Refinement of the PA0803-Pyocyanin Complex

As the cell parameters of the PA0803-pyocyanin complex crystal are very similar to the ones of the apo crystal (Yu, 2009), this structure could be used in rigid body refinement to determine the structure of the PA0803 pyocyanin complex.

Table 35: Refinement statistics for *P. aeruginosa* PA0803 in complex with pyocyanin (PYO). Values in parentheses are for the highest resolution shell.

Dataset	PA0803 + PYO
Resolution range (Å)	46.55-1.62 (1.65-1.62)
R_{work} (%)	17.6 (35.5)
R_{free} (%)	20.2 (38.5)
No. of non-H atoms	
Protein	4430
Ion	-
Ligand	52
Water	705
R.m.s. deviations	
Bonds (Å)	0.005
Angles (°)	0.999
Average B factors (Å^2)	
Protein	27
Ion	-
Ligand	53
Water	40
Ramachandran plot	
Favored regions (%)	98.1
Outliers (%)	0
MolProbity score#	1.19
PDB entry code	-

#As reported by *MolProbity* (Chen *et al.*, 2010)

After the initial rigid body refinement cycle in *REFMAC5*, the structural model of PA0803 was correctly placed in the electron density, confirming the presence of four PA0803 molecules in the asymmetric unit. Additional positive $m|F_{obs}|-D|F_{calc}|$ electron density clearly demonstrated the presence of pyocyanin in all four binding sites present in the asymmetric unit. The model was improved manually and refined using *phenix.refine* (Table 35). The structure of the PA0803 pyocyanin complex was finally refined to an R_{work} = 17.6% and an R_{free} = 20.2% with a *MolProbity* score of 1.19. In the PA0880 structure, 98.1% of the residues are in Ramachandran-favored regions.

Additionally, the diffraction data obtained from PA0803, PA1353 and PA4641 crystals in previous studies (Yu, 2009; Kalawy-Fansa, 2010) were reprocessed with *XDS* and *Aimless* (Evans, 2006, 2011; Kabsch, 2010) and the structures were refined to lower R-factors using *phenix.refine* (Afonine *et al.*, 2012). The statistics can be found in the appendix.

6.5.3 Crystal Structure of PA0803 in Complex with Pyocyanin

The crystal structure of PA0803 in its apo form as well as in complex with pyocyanin at 25% occupancy have been extensively discussed in a previous work (Yu, 2009). For this reason, only a short overview about the binding mode of pyocyanin to PA0803 in the newly determined structure with 100% pyocyanin occupancy will be given here.

In the crystal structure of PA0803 in complex with pyocyanin (Figure 43), the phenazine compound can be observed in all four binding pockets of PA0803 in the asymmetric unit by positive $m|F_{obs}|-D|F_{calc}|$ electron density. The main binding interactions are provided by H49 and Y122* (* indicating that the residue originates from the other chain of the homodimer) which are both 3.5 Å apart from the pyocyanin molecule. Pyocyanin is sandwiched between these two side chains by π-stacking interactions. Additionally, Y93* and Y142* provide further hydrophobic interactions to position pyocyanin in the binding pocket. Two salt bridges between the hydroxyl group of Y142* and the carboxyl group of E51 as well as the latter and the imidazole ring of H49 stabilize the binding pocket. The orientation of pyocyanin in the PA0803 binding pocket could easily be determined by the position of its hydroxyl group, which is clearly visible in the electron density. The hydroxyl group is pointing away from the aromatic residues Y93*, Y122* and Y142, towards the more open site of the binding pocket. Interestingly, this pyocyanin orientation is different from the one observed in the structure solved in a previous study (Yu, 2009). This difference in orientation might be explained by ambiguous electron density caused by the reduced occupancy of pyocyanin in that structure.

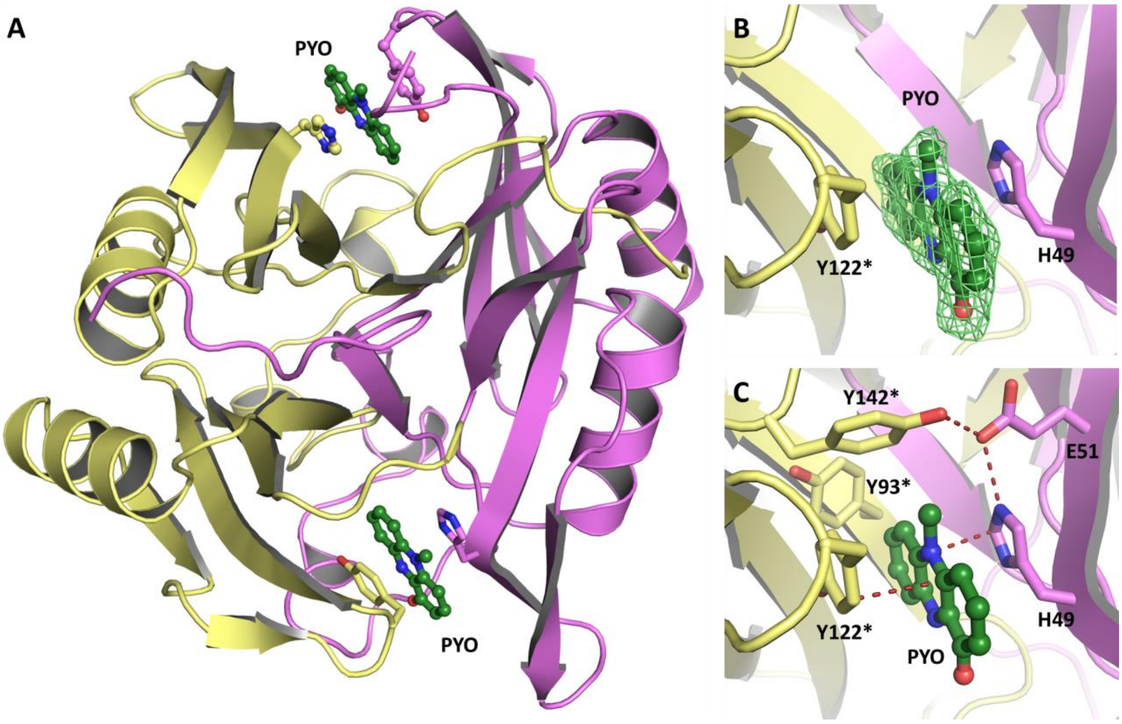


Figure 43: Crystal structure of the PA0803 pyocyanin complex. A: The pyocyanin binding sites of the PA0803 homodimer are formed at the dimer interfaces. Pyocyanin is sandwiched between two aromatic residues. B: Positive m|F_{obs}|-D|F_{calc}| density (σ-level = 2) confirms the presence of pyocyanin. Its orientation is determined by the position of its hydroxyl group. C: Binding interactions of pyocyanin and PA0803 are mainly established by π-stacking between H49 and Y122*. Y93* and Y142* provide further π-interactions. Salt bridges between Y142* and E51 as well as H49 and E51 further stabilize the pyocyanin binding pocket.

The dissociation constant K_D of the interaction between pyocyanin and PA0803 has been previously determined as 3.92 ± 0.22 µM for PA0803 (Yu, 2009). PA0803, PA1353 and PA4641 may act as binding proteins for pyocyanin that should transport the compound to efflux pumps. These efflux pumps in turn may secrete pyocyanin into the extracellular space. For this reason, the moderate interaction between pyocyanin and its binding proteins might be of physiological importance.

6.5.4 ITC Experiments with PA0803, PA1353 and PA4641

As previously mentioned, PA0803, PA1353 and PA464 all bind pyocyanin in the low micromolar range (Yu, 2009; Kalawy-Fansa, 2010). In order to investigate the specificity of PA0803, PA1353 and PA4641 towards pyocyanin, phenazines in general and towards other compounds containing a number of (aromatic) rings different from three (as the phenazines), ITC experiments were conducted with these proteins and a number of substances (Figure 44).

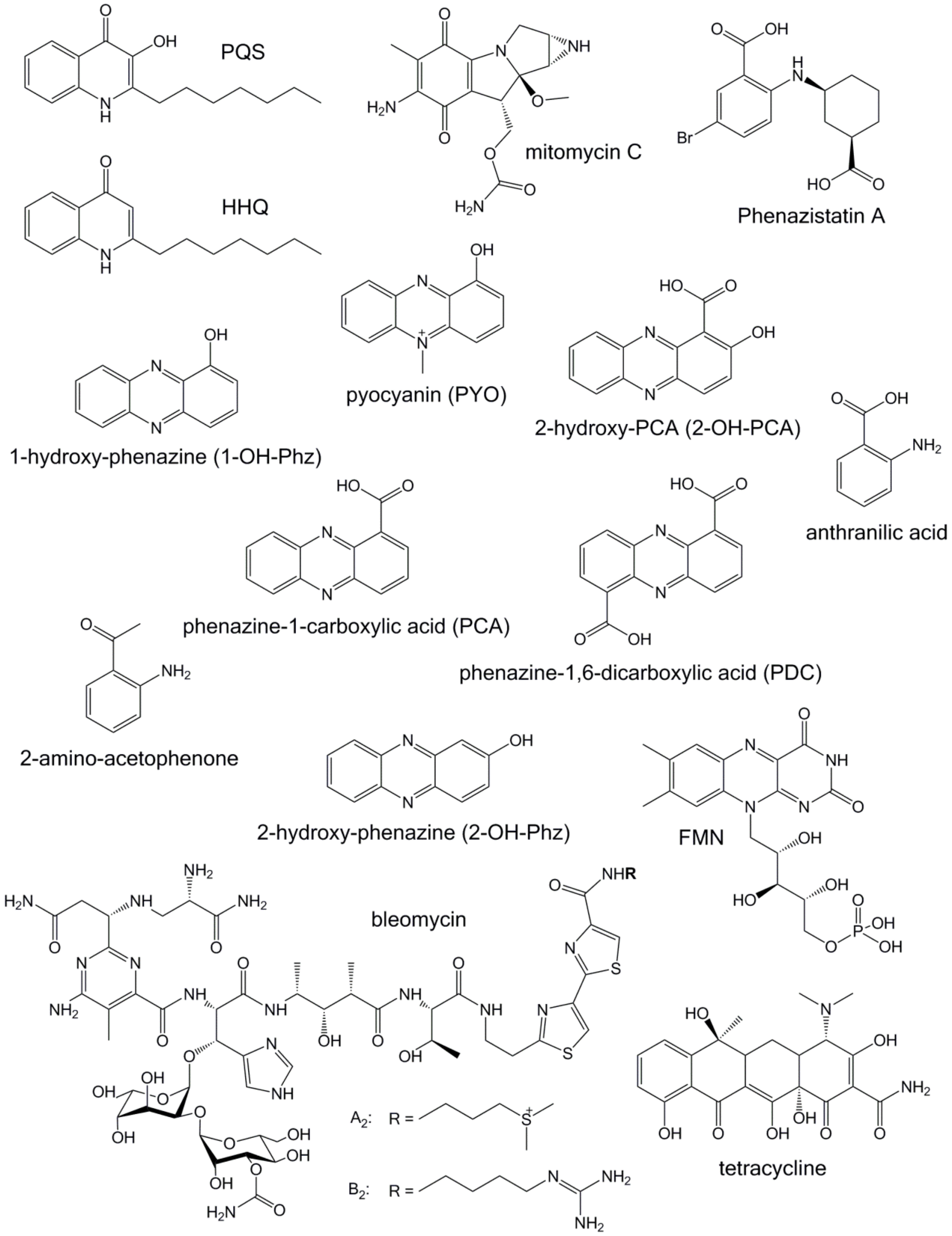


Figure 44: Compounds assayed for binding to PA0803, PA1353 and PA4641 in ITC experiments. Bleomycin, a complex glycopeptide antibiotic was also assayed for binding. FMN = Flavin mononucleotide, HHQ = 2-heptyl-4-quinolone, PQS = *Pseudomonas* Quinolone Signal (2-heptyl-3-hydroxy-4-quinolone).

The ability of PA0803, PA1353 and PA4641 to bind some of these compounds has also been investigated previously (Yu, 2009; Kalawy-Fansa, 2010). The compounds tested in ITC experiments with PA0803, PA1353 and PA4641 (Table 36) were chosen according to their relevance in *P. aeruginosa* physiology, or their possible presence in habitats of *P. aeruginosa*. For instance, 2-amino-acetophenone was chosen because it gives the characteristic smell to *P. aeruginosa*. Phenazistatin A is a potent inhibitor of phenazine biosynthesis and a drug lead that targets PhzB with an affinity of K_D = 51 nM (M. Mentel, unpublished data). HHQ and PQS were chosen because they are involved in *P. aeruginosa* QS which controls phenazine biosynthesis. Apart from a variety of phenazine compounds, also three antibiotics, mitomycin C, tetracycline and bleomycin, an anti-cancer drug, were chosen for the ITC binding assay.

Table 36: ITC experiments with potential pyocyanin resistance factors PA0803, PA1353 and PA4641. $^+$ = tested in a previous study by Yu, 2009. * = tested in a previous study by Kalawy-Fansa, 2010. - = no binding detected. Comparable K_D's of PA0803 and pyocyanin indicate the reproducibility of the results. K_D's obtained from this study are calculated from at least three independent measurements. Phz = phenazine.

ligand	PA0803	PA1353	PA4641
anthranilic acid	-*	-*	-*
2-amino-acetophenone	-	-	-
phenazistatin A	-	-	-
HHQ	-*	-*	-*
PQS	-	-	-
1-OH-Phz	75.7 ± 7.2 µM*	4.7 ± 0.68 µM*	0.8 ± 0.18 µM*
2-OH-Phz	-	0.27 ± 0.5 µM	0.14 ± 0.02 µM
PYO	3.92 ± 0.22 µM$^+$ 4.4 ± 0.32 µM* 2.63 ± 0.15 µM	3.4 ± 0.25 µM* 0.99 ± 0.19 µM	0.64 ± 0.35 µM* 0.40 ± 0.11 µM
PCA	-* -	24.1 ± 3.1 µM* 9.1 ± 3.0 µM	8.9 ± 0.7 µM* 7.7 ± 0.31 µM
2-OH-PCA	-	-	-
PDC	-	-	-
mitomycin C	-	-	-
FMN	-*	14.1 ± 2.6 µM*	5.9 ± 1.2 µM*
tetracycline	-	-	-
bleomycin	4.94 ± 0.59 µM	2.5 ± 0.74 µM	3.27 ± 0.35 µM

In the ITC experiments conducted in this and previous studies (Yu, 2009; Kalawy-Fansa, 2010), none of the three proteins PA0803, PA1353 or PA4641 shows affinity for anthranilic acid, 2-amino-acetophenone, HHQ or PQS, meaning compounds possessing only one or two aromatic rings. Fortunately, no binding affinity is detected for phenazistatin A, a promising phenazine biosynthesis inhibitor. This effect is promising for drug development against phenazine biosynthesis, as the potential resistance proteins do not interfere with this compound. In a previous study, FMN was demonstrated to bind to PA1353 and PA4641 with K_Ds of 14.1 ± 2.6 µM and 5.9 ± 1.2 µM, respectively, but no affinity of FMN to PA0803 was detected (Kalawy-Fansa, 2010). Tetracycline, an antibiotic containing four 6-rings, has no affinity to any of the three proteins PA0803, PA1353 and PA4641. The scenario is the same for mitomycin C, an antibiotic produced by *Streptomyces* spp. Interestingly, the complex antibiotic bleomycin was demonstrated to bind to PA0803, PA1353 and PA4641 (Figure 45) with affinities of 5.71 ± 0.15 µM, 2.5 ± 0.74 µM and 3.27 ± 0.35 µM respectively. However, the proteins slightly precipitate in the ITC machine upon bleomycin binding, indicating aggregation of the mixture. Still, they could have a similar function as another member of the glyoxalase superfamily, BLMT, a bleomycin-binding protein from *S. verticillus* (PDB entry 1JIF, Sugiyama *et al.*, 2002).

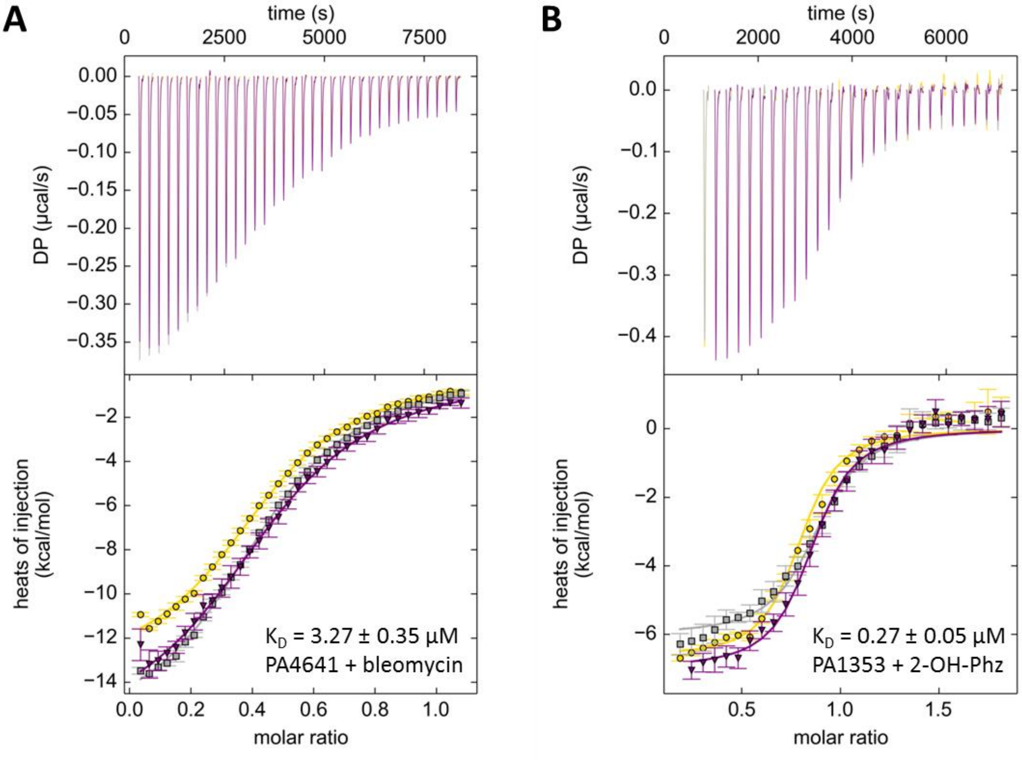


Figure 45: Representative ITC experiments. A: Titration of bleomycin into PA4641 is followed by binding of bleomycin to PA4641 with a K_D = 3.27 ± 0.35 µM. However, the protein seems to precipitate in the ITC upon bleomycin binding. B: Titration of 2-OH-Phz into PA1353 yields binding of 2-OH-Phz to PA1353 with a K_D = 0.27 ± 0.05 µM.

The class of compounds that are preferentially bound by PA0803, PA1353 and PA4641 are the phenazines, as demonstrated also in earlier studies. In this study, binding of pyocyanin to PA0803, PA1353 and PA4641 and of PCA to PA1353 and PA4641 was confirmed (Table 36). Additionally, binding of 2-OH-Phz to PA1353 and PA4641 with affinities of K_D = 0.27 ± 0.5 µM and 0.14 ± 0.02 µM was demonstrated, respectively. However, no binding of 2-OH-PCA or PDC was observed for any of the three proteins. A reason for this might be that the two carboxyl groups of PDC are too bulky for the binding pockets of PA0803, PA1353 and PA4641. Taken together, pyocyanin is the best-binding phenazine to the proteins PA0803, PA1353 and PA4641, apart from 2-OH-Phz that binds better to PA1353 and PA4641. PA4641 has the highest affinity for phenazines in general. All three proteins are most likely preferentially binding compounds with three condensed aromatic moieties, but compounds other than phenazines may not be generally excluded as seen for bleomycin.

6.5.5 Spot Plate Assays with PA0803, PA1353 and PA4641

Spot plate assays using both *E. coli* and *P. aeruginosa* were performed in order to investigate the potential role of PA0803, PA1353 and PA464 to mediate pyocyanin resistance *in vivo*. For this purpose, *E. coli* carrying a pET19m plasmid encoding PA0803 or PA1353 or PA3127 or PA4641 were spotted on pyocyanin-doted agar plates (Figure 46). For *P. aeruginosa*, ΔPA0803 or ΔPA1353 or ΔPA4641 transposon mutants of PAO1 or PA14 strains were spotted on pyocyanin-doted agar, or *P. aeruginosa* PAO1 were transformed with a pHERD30T plasmid encoding either PA0803 or PA1353 or PA4641. The latter approach resulted in *gain-of-function* strains which were tested on pyocyanin-agar plates as well.

Both the spot plate experiments with *E. coli* or *P. aeruginosa* were not successful. First, *E. coli* was very sensitive to pyocyanin and grew on remarkably lower pyocyanin concentrations (10 µM) compared to *P. aeruginosa* (> 1 mM). This is in agreement with earlier findings (Hassett *et al.*, 1992). *E. coli* producing no protein or PA4641 still grew when spotted at a density of $5\text{x}10^5$ cells/mL, although PA4641 shows the highest binding affinity for pyocyanin *in vitro*. *E. coli* producing either PA0803 or PA3127 (the protein has no affinity for pyocyanin) still grew when spotted at a density of $5\text{x}10^4$ cells/mL, while *E. coli* producing PA1353 still grew when spotted at a density of 10^4 cells/mL. However, these differences are not significant, as they usually range between a 10^3- to 10^6-fold change in growth.

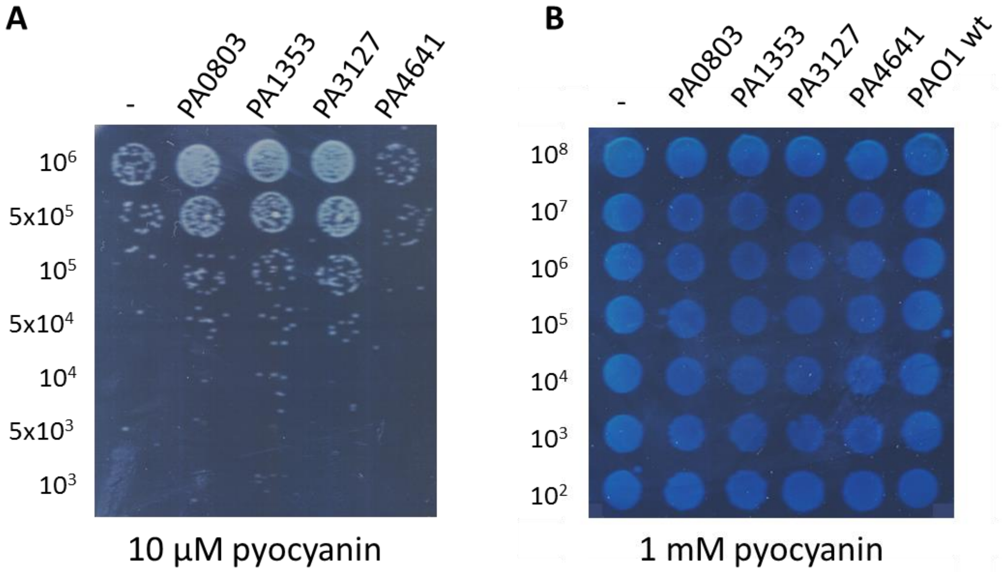


Figure 46: Pyocyanin spot plate assays. A: Spot plate assay of *E. coli* BL21(DE3) transformed with pET19m encoding either no protein (-) or PA0803, PA1353, PA3127 or PA4641. No significant evidence was found for one of the proteins to mediate pyocyanin resistance, although a slight effect is visible for PA1353. B: Spot plate assay of *P. aeruginosa* PAO1 *gain-of-function* strains transformed with pHERD30T encoding either no protein (-) or PA0803, PA1353, PA3127 or PA4641. Even at the maximum possible concentration of pyocyanin in aqueous solutions or media (1 mM), low concentrations of *P. aeruginosa* can still grow on pyocyanin doted agar. Equal results are obtained at even lower bacterial concentrations (up to 10^{-5} cells/mL) and with PAO1 and PA14 transposon mutants (data not shown).

As pyocyanin is producing ROS, *E. coli* are probably under high oxidative stress, which in turn inhibits normal growth and kills the bacteria. This effect is possibly stronger than the chaperone-like effect of PA0803, PA1353 or PA4641. Another difficulty could be the lack of appropriate efflux pumps that may act together with PA0803, PA1353 or PA4641 to export pyocyanin to the extracellular space. These facts render it more difficult to investigate the function of PA0803, PA1353 and PA4641.

For *P. aeruginosa* spot plates, a maximum concentration of 1 mM pyocyanin was used, which is the highest possible concentration of pyocyanin in aqueous solutions or media. However, *P. aeruginosa* cells carrying pHERD30T plasmids, each encoding either no protein or PA0803, PA1353, PA3127 or PA4641 still grew on spot plates containing 1 mM pyocyanin when spotted at a density of 10^2 cells/mL. The cells were even diluted to 10^{-5} cells/mL, but still *P. aeruginosa* was growing. Exactly the same results were obtained with PAO1 and PA14 transposon mutants ΔPA0803, ΔPA1353, ΔPA3127 and ΔPA4641.

There are several problems encountered when investigating the potential function of PA0803, PA1353 and PA4641 in their natural origin organism *P. aeruginosa*. First, *P. ae-*

ruginosa possess several general anti-oxidative defense mechanisms, which are upregulated with increasing pyocyanin concentration (Hassett *et al.*, 1992). These defense mechanisms probably interfere with the chaperone-like function of PA0803, PA1353 and PA4641. Second, when investigating the effect of the absence of PA0803 in a *P. aeruginosa* Δ0803 transposon mutant, PA1353 and PA4641 are still present and might complement the function of PA0803. For this reason, no change in phenotype of *P. aeruginosa* can ultimately be observed. Attempts to generate double and triple deletion mutants of PA0803, PA1353 and PA4641 failed during this study due to difficulties in the overlap PCR step. Third, pyocyanin is a terminal signaling factor in the *P. aeruginosa* QS system (Dietrich *et al.*, 2006). Pyocyanin regulates its own *PYO Stimulon*, including 51 genes, 22 of them upregulated and 29 of them down regulated. Among the upregulated genes, eight genes encode putative efflux pumps, which might transport pyocanin to the extracellular space, and two encode transcriptional regulators (Dietrich *et al.*, 2006), for instance the transcriptional regulator SoxR and its regulon. The expression of SoxR and the efflux pump MexGHI-OpmD, which is encoded in the *phz2* operon of *P. aeruginosa*, is upregulated by 30-fold upon exposure to pyocyanin. This response to pyocyanin is detectable after 10 min of exposure (Dietrich *et al.*, 2006). Taken together, exposure of *P. aeruginosa* to pyocyanin subjects dramatic effects to the gene expression profile of the bacterium, activating several hierarchical levels of genes in a very short time frame. Finally, pyocyanin production itself is regulated by the QS systems of *P. aeruginosa*, which are able to sense the pyocyanin concentration and to dramatically change the expression profile of *P. aeruginosa*, and hence the phenotype and behavior of the bacterium. For these reasons, it is very hard to study the potential function of PA0803, PA1353 and PA4641 in *P. aeruginosa*.

6.6 Discovery of Bleomycin Resistance in *P. aeruginosa* PAO1

PA3127 is a 29.6 kDa protein with calculated extinction coefficient $\varepsilon = 43555\ M^{-1}\ cm^{-1}$ and a theoretical pI of 5.3. Interestingly, the molecular weight of this βαβββ-module protein is about double that of usual two-module containing βαβββ-module proteins. *Phyre*^{2} predictions reveal that PA3127 is indeed a unique βαβββ-module resistance protein in *P. aeruginosa*, because it carries an additional N-acetyltransferase domain. The presence of this domain indicates that this protein fulfils a particular function, probably mediating resistance against a certain antibiotic in the pathogen. This view is encouraged by the fact that a similar domain is found in a bleomycin N-acetyltransferase (BAT) produced by *S. verticillus* (Oda *et al.*, 2010). Possibly, PA3127 is the BAT of *P. aeruginosa*.

6.6.1 Background - Bleomycin and Bleomycin Resistance Proteins

The natural glycopeptide antibiotic bleomycin is on the World Health Organization's List of essential medicines that are needed for a basic health system. Bleomycin is produced by *Streptomyces* spp. and the term bleomycin therefore includes a number of structurally related compounds. For instance, bleomycin is referred to as a mixture of bleomycin A_2 and B_2 in cancer therapy (Figure 47), where it is to treat Hodgkin's lymphoma, squamous cell carcinoma and testicular cancer, but also plantal warts or pleurodesis.

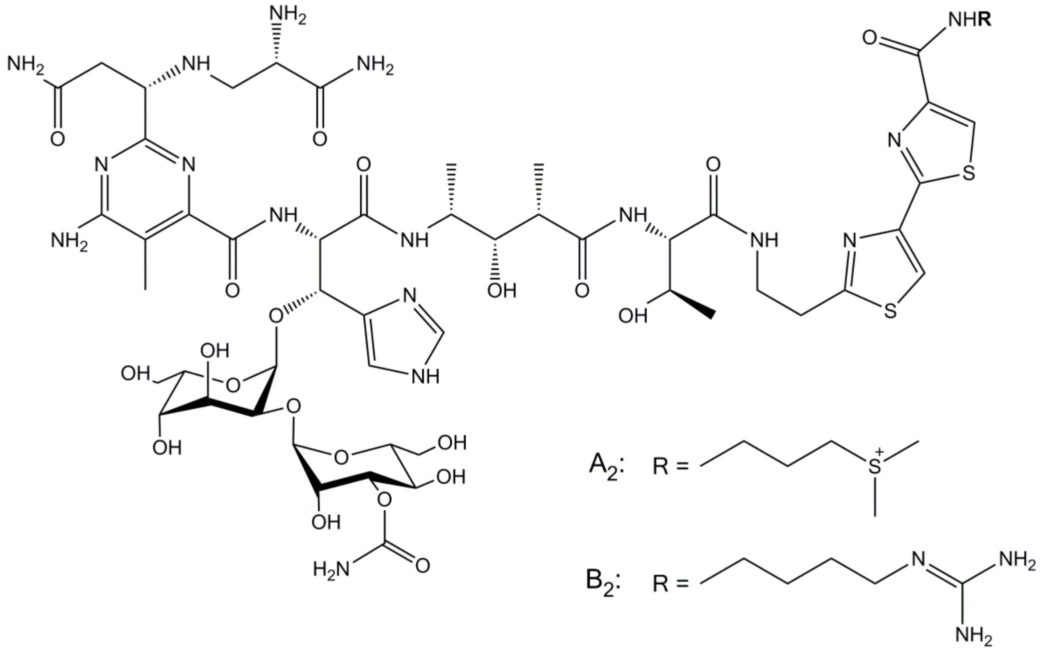


Figure 47: The complex glycopeptide antibiotic bleomycin. The anti-cancer drug bleomycin (for instance Bleomedac® used in this study) is a mixture of bleomycin A_2 and B_2. Bleomycin is biosynthesized by non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) in *Streptomyces* spp.

The metal-bound form of bleomycin causes DNA scission in cells that are not resistant to the antibiotic. Two different mechanisms of action of bleomycin are discussed. On the one hand, bleomycin might first chelate Fe(II), thereby producing a pseudoenzyme which reacts with oxygen. In this manner, superoxide and hydroxide radicals are produced, that in turn cleave the DNA. On the other hand, bleomycin might first bind to the DNA and then induce scission by abstracting a hydrogen atom from a base. This hydrogen abstraction will result in DNA strand cleavage (Hecht, 1999).

In order to achieve resistance against self-produced bleomycin, *S. verticillus* produces two types of resistance proteins: BLMT, a bleomycin binding protein, and BAT, a bleomycin-N-acetyltransferase (Figure 48).

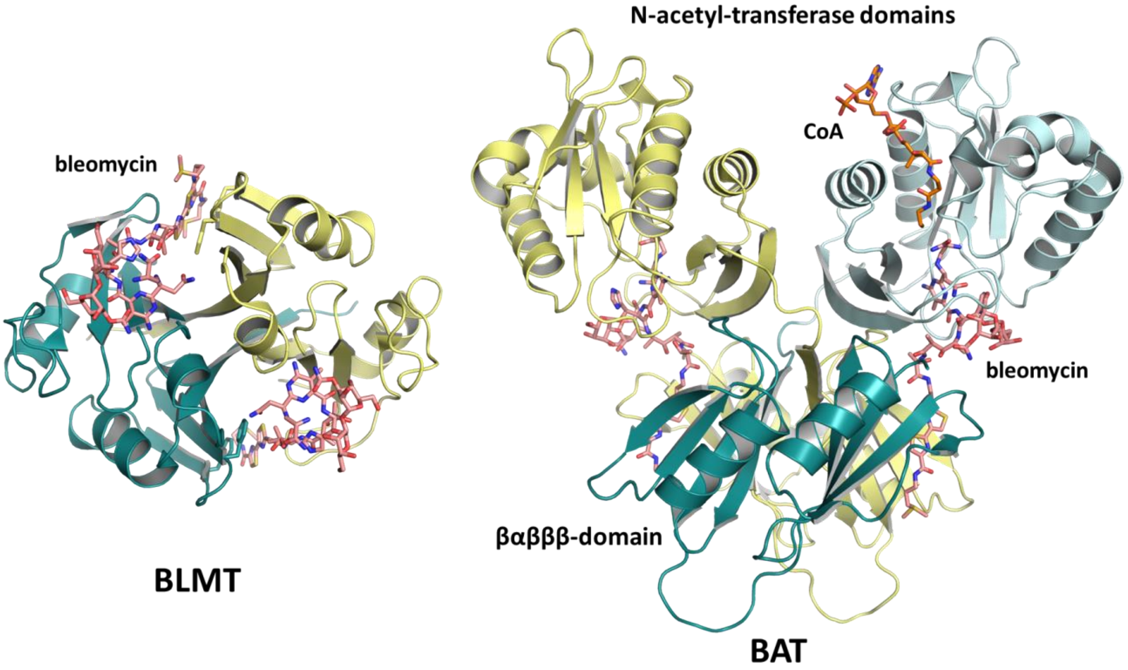


Figure 48: Bleomycin resistance proteins from *S. verticillus*. BLMT (PDB entry 1JIF, Sugiyama *et al.*, 2002) is a bleomycin binding protein. BLMT is a βαβββ-module protein and acts as a chaperone towards the metal-bound form of bleomycin. BLMT does not alter bleomycin chemically. BAT (PDB entry 2ZW7, (Oda *et al.*, 2010) is a bleomycin N-acetyltransferase that catalyzes the acetylation of the primary amine of metal-free bleomycin under AcCoA consumption.

BLMT is a homodimer and acts like a chaperone towards the metal-bound form of bleomycin, and does not alter the compound chemically. The protein belongs to the βαβββ-module proteins and the family of glyoxalase II proteins (PF12681). BLMT possesses two bleomycin binding sites at its dimer interfaces. In contrast, BAT is an enzyme that uses AcCoA to acetylate the primary amine of metal-free bleomycin, thereby rendering it inactive. BAT of *S. verticillus* forms a homodimers via dimerization of the C-terminal βαβββ-module domain. The N-terminal N-acetyltransferase domain is monomeric and binds a CoA molecule (the product of the acetylation reaction). Bleomycin is bound in both domains and stretches across the whole protein. The bis-thiazole moiety of bleomycin is bound in the βαβββ-module domain while the primary amine of bleomycin is in proximity of the potential acetyl group.

Interestingly, the domain architecture of the *S. verticillus* BAT differs from the architecture of the potential BAT PA3127 from *P. aeruginosa* (Figure 49). The difference in domain arrangement indicates that BAT and PA3127 probably do not have a common ancestor, but are derived from different gene editing events.

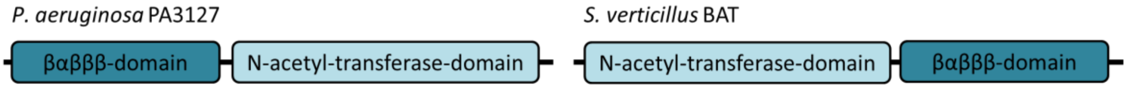


Figure 49: Differences in domain arrangement of *P. aeruginosa* PA3127 and *S. verticillus* BAT.

The difference in domain arrangement will not allow MR of PA3127 crystals with full length BAT from *S. verticillus*, but rather require MR using separate domains of the protein or other βαβββ-module domain or N-acetyltransferase domains as MR models.

6.6.2 Evidence for PA3127 Bleomycin and AcCoA Binding and Catalytic Acetyltransferase Activity

In order to test the hypothesis that PA3127 is a bleomycin N-acetyltransferase, the protein was subjected to ITC experiments and an HPLC-MS activity tests. In the ITC experiments, PA3127 was tested for binding towards AcCoA, CoA and bleomycin as well as some of the substances that were tested in the experiments with the pyocyanin binding proteins PA0803, PA1353 and PA4641, namely pyocyanin, PCA, 2-amino-acetophenone and phenazistatin A (Figure 44). However, pyocyanin, PCA, 2-amino-acetophenone and phenazistatin A had no affinity towards PA3127 or the N-terminal βαβββ-module domain of PA3127 alone. In contrast, AcCoA, CoA and bleomycin demonstrate to have high affinity to PA3127 (Figure 50). PA3127 bind its product CoA with an affinity of K_D=2.7 ± 0.32 µM (data not shown), while it binds its cosubstrate AcCoA with a 10-fold higher affinity of K_D = 529 ± 33 nM (Figure 50A). The affinity of PA3127 towards bleomycin is in the high nanomolar range with K_D = 413 ± 4 nM (Figure 50B).

The data indicate that PA3127 might indeed be a bleomycin N-acetyltransferase. The high affinities of PA3127 for these compounds are probably needed for the physiological function of PA3127, as the enzyme has to detect, bind and detoxify already very small amounts of bleomycin, thereby preventing the cell from DNA damage. In order to ensure supply with AcCoA, the affinity for the cosubstrate is probably high, so that immediate acetylation of bleomycin can occur upon uptake of this toxic compound.

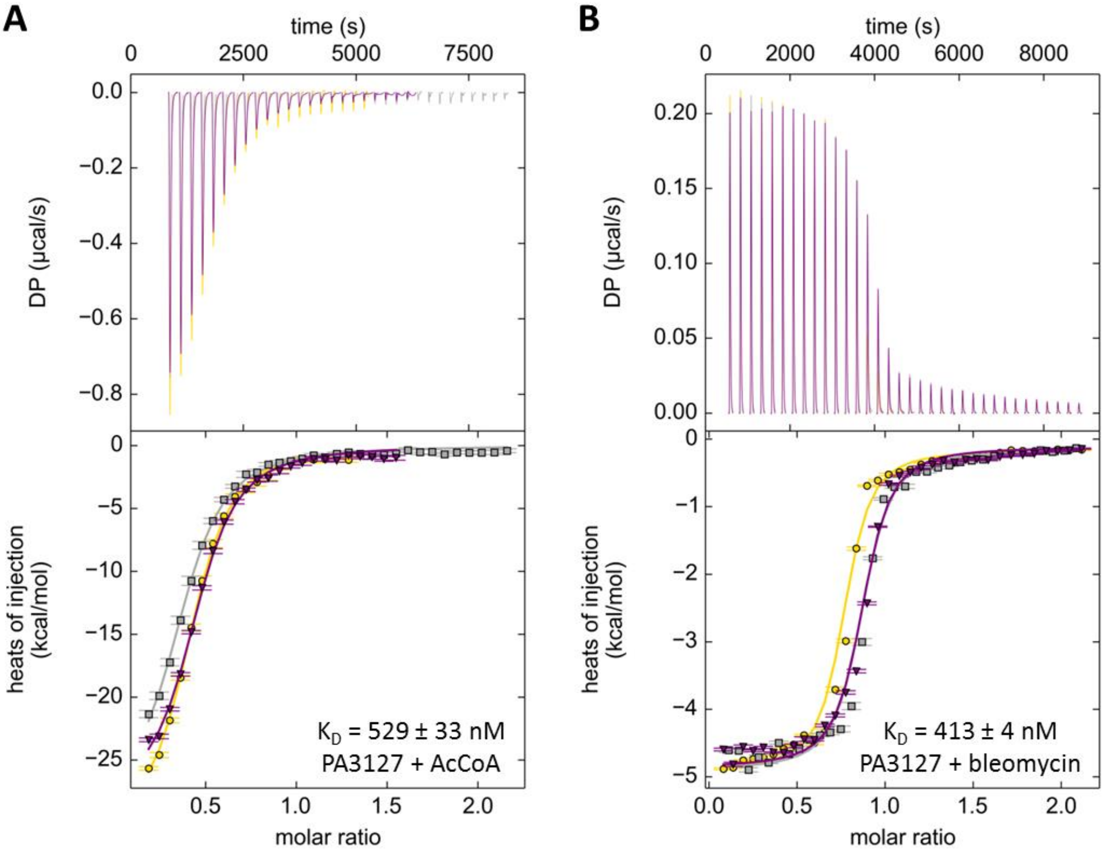


Figure 50: ITC experiments with PA3127. Both binding events are exothermic and occur in 1:1 stoichiometry. The ITC experiments were performed in two different ITC machines, which causes differences in the plots. A: Binding of the cosubstrate AcCoA to PA3127 occurs with a K_D of 529 ± 33 nM. B: Binding of the antibiotic substrate bleomycin to PA3127 occurs with a K_D of 413 ± 4 nM.

In the activity test, PA3127 was assayed together with AcCoA, NADPH, $NADP^+$, bleomycin and PA3128, the protein encoded in an operon together with PA3127, in 50 mM BisTris pH 5.8, 6.5 or 7.2. However, it was not possible to detect AcCoA/CoA and bleomycin/acetyl-bleomycin simultaneously under the conditions tested in HPLC elution. In these elution profiles, two reaction times (5 min and 1 or 2 h) are compared for each buffer condition (Figure 51). In each sample, a clear AcCoA peak was detectable. In contrast, a significant CoA peak is only detectable at a physiological pH of 7.2. In the control experiment (no enzymes) no CoA peak was detected.

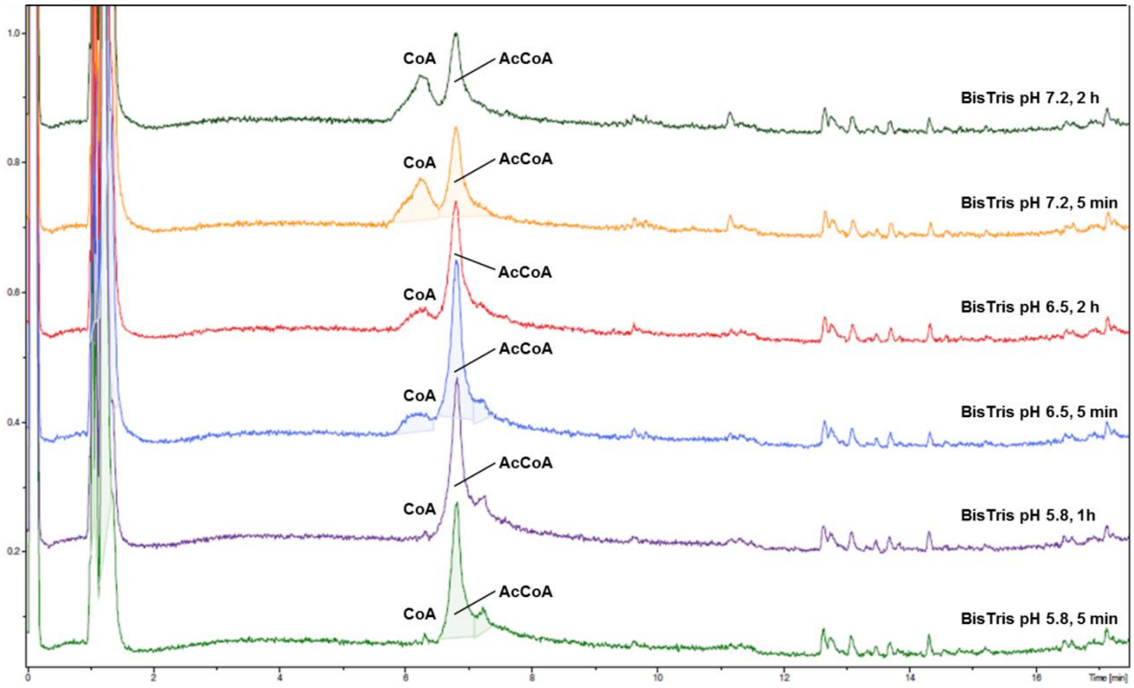


Figure 51: HPLC elution profiles of the reaction mixtures from the PA3127 bleomycin N-acetyltransferase-assay. The shortest and longest incubation time of each buffer condition are given. A clear AcCoA peak can be detected in all samples, however, CoA only seems to appear in significant amount at a physiological pH of 7.2. At pH 6.5, only little amounts of CoA are produced, while no CoA is produced at pH 5.8. This indicates that PA3127 utilizes AcCoA and produces CoA, and is indeed an acetyltransferase. However, the substrate bleomycin or the reaction product acetyl-bleomycin could not be detected in this experimental setup.

No differences in the AcCoA/CoA peaks in the samples with different incubation time from one buffer condition are observed, meaning that the reaction had taken place after 5 min. Possibly, no further conversion of AcCoA is observed because bleomycin is completely consumed. These experiments indicate that PA3127 is able to use AcCoA in an acetyltransferase reaction to form CoA.

6.6.3 Crystallization and Data Collection of PA3127

PA3127 was screened at various concentrations without or with bleomycin, AcCoA or CoA and combinations of those, using various crystallization screens. Only one condition yielded rod-shaped crystals (JCSG Core III F9) (Bittner, 2012), which were optimized. The crystals were rather solid, but difficult to reproduce. A first well-diffracting crystal was obtained from 0.1 M MES pH 5.6, 0.8 M ammonium sulfate (Figure 52). Surprisingly, a better-diffracting crystal could be obtained after letting it dehydrate in a 24-well plate for 377 days.

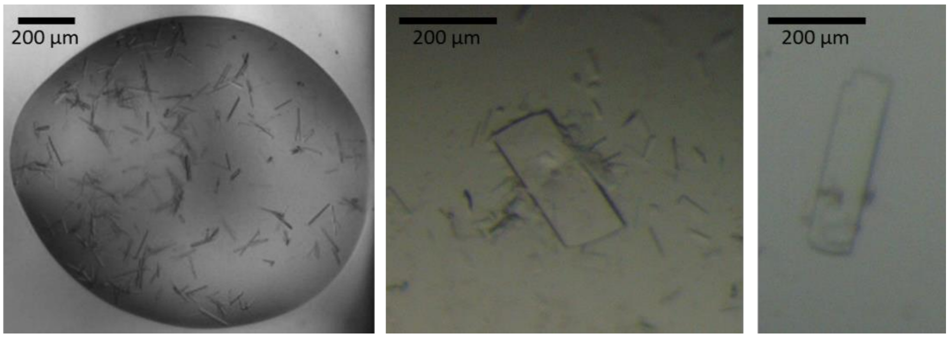


Figure 52: Crystals of *P. aeruginosa* PA3127. Initial hit in JCSG Core III F (left) and optimized crystals from a 24-well plate (middle/right).

The PA3127 crystal was cryoprotected with 25% (*v/v*) glycerol before diffraction data collection on beamline 14.1 at BESSY II Synchrotron of the HZB, Berlin, Germany (Table 37). A dataset of the standard cell protein crystal was collected in the course of a bachelor thesis under supervision of the author (Bittner, 2012) as 100 images of 1° rotation on a MarCCD Chess detector at a wavelength of λ = 0.9184 Å.

Table 37: Data collection statistics for *P. aeruginosa* PA3127 in complex with residual (Ac)CoA. Values in parentheses are for the highest resolution shell.

Dataset	PA3127 – standard cell	PA3127 – dehydrated cell
Wavelength (Å) / beamline[≠]	0.9184 / BESSY II, BL14.1	0.9184 / BESSY II, BL14.1
Resolution range (Å)	96.58-2.95 (3.04-2.95)	49.03-2.48 (2.64-2.48)
Space group	$P2_1$	$P2_1$
Unit cell parameters (Å)	106.9 121.0 116.2	97.6 120.2 106.8
(°)	90.0 115.35 90.0	90.0 112.8 90.0
Mosaicity (°)[†]	0.102	0.105
Total No. of measured reflections	121303 (9989)	259220 (17051)
Unique reflections	55026 (4540)	70799 (4522)
Multiplicity	2.2 (2.2)	3.7 (3.8)
Mean I/σ(I)	8.8 (1.5)	15.7 (2.0)
Completeness (%)	97.7 (98.5)	99.1 (98.8)
R_{meas} (%)	10.6 (69.6)	7.4 (88.1)
$R_{p.i.m.}$ (%)	6.8 (44.5)	3.8 (41.3)

[†] Mosaicity values reported by *XDS* (Kabsch, 2010)

Diffraction data of the dehydrated PA3127 were collected as 1900 images of 0.1° rotation using a PILATUS 6M pixel detector at a wavelength of λ = 0.9184 Å. The differences in cell parameters between the crystal with the standard cell and the dehydrated cell are obvious, as the cell axes *a*, *b* and *c* shrink by 9.3 Å, 0.8 Å and 9.4 Å, respectively. However, no solution of the PA3127 crystal structure could be identified using MR approaches with different models or the MR engine *BALBES* (Long *et al.*, 2007). Cocrystallization of with 5-amino-2,4,6-triiodoisophthalic acid (I3C) failed to produce crystals, while selenomethionine labeling and heavy atom soaking produced crystals of low quality, either with weak or anisotropic diffraction properties.

6.6.4 *Divide and Conquer* Approach

Finally, the *Divide and Conquer* approach was applied to the hypothetical bleomycin-N-acetyltransferase PA3127. For this purpose, separate constructs of the domains were designed according to structure predictions by *Phyre*^{2} (Kelley & Sternberg, 2009) in a study parallel to this work and under supervision of the author (Popp, 2013). Diffracting crystals of both the βαβββ-domain PA3127_1-114 (PA3127-NTD) and the N-acetyltransferase domain PA3127_123-267 (PA3127-CTD) were obtained. However, it was not possible to successfully determine the PA3127-CTD structure using a 2.7 Å dataset (Popp, 2013). This was probably due to low data quality. In contrast, the PA3127-NTD structure was determined to a resolution of 1.55 Å (Figure 53, Popp, 2013, and Table 48 & Table 49 in the appendix).

The N-terminal βαβββ-domain of PA3127, PA3127-NTD, folds into two βαβββ modules and forms a homodimer. Binding of the bis-thiazole moiety of bleomycin probably involves residues Y31, Y34 and W96, as seen in BLMT from *S. verticillus*. W96 is located in a flexible loop which probably has to undergo a conformational change upon bleomycin binding. A partial MR solution for the PA3127 crystal was obtained in *phenix.phaser* using the structure of PA3127-NTD. This solution includes eight PA3127-NTD molecules (four PA3127-NTD homodimers) in the asymmetric unit. Crystal packing suggested that there is enough space for the missing eight N-acetyltransferase domains. However, no full MR solution was obtained by using the partial solution with various N-acetyltransferase structures deposited in the PDB. Hence, it was crucial to determine the structure of the PA3127 N-acetyltransferase domain PA3127-CTD alone.

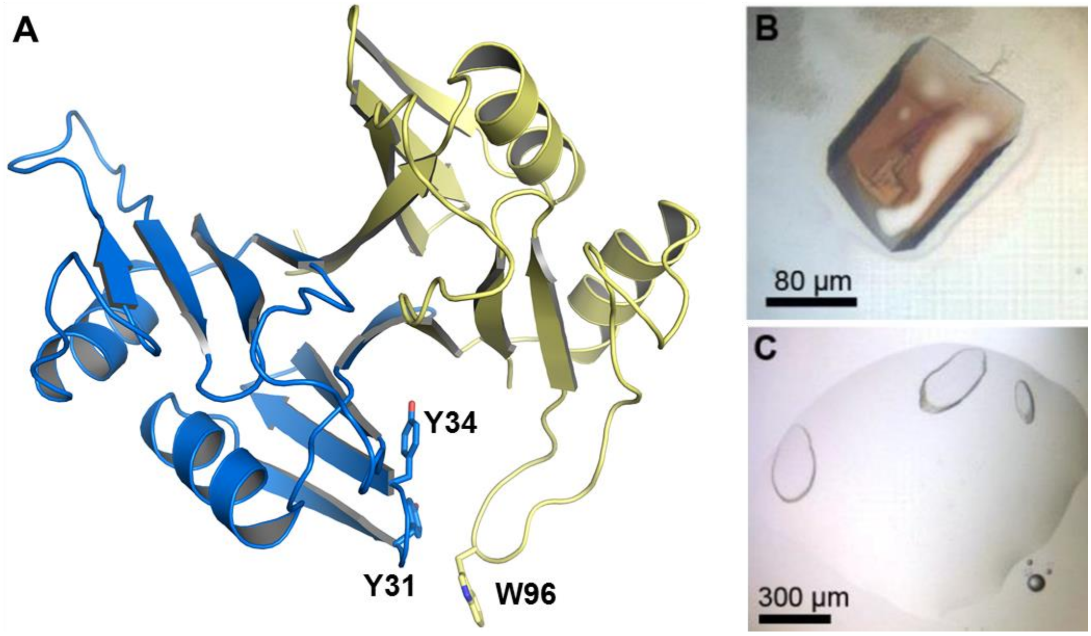


Figure 53: Results of the *Divide and Conquer* approach obtained in a study under supervision of the author (Popp, 2013). A: Crystal structure of PA3127-NTD. PA3127 is a βαβββ-module protein and forms a homodimer in solution. Residues Y31, Y34 and W96 are probably involved in binding the bis-thiazole moiety of bleomycin. B: Optimized crystal of PA3127-NTD. C: Optimized crystals of PA3127-CTD. Figure adapted from Popp, 2013.

6.6.5 Crystallization and Diffraction Data Collection of PA3127-CTD

PA3127-CTD was screened again in the presence as well as in the absence of 2 mM AcCoA during this study. The only initial hit was found in condition A1 of the MIDAS screen (0.1 M HEPES/NaOH pH 6, 50% (*v/v*) polypropylene glycol 400, 5% (*v/v*) DMSO). However, these crystals (Figure 54) could not be reproduced in a 24-well format. For this reason, PA3127-CTD crystals were isolated directly from this crystallization experiment and were tested for diffraction without further cryoprotection.

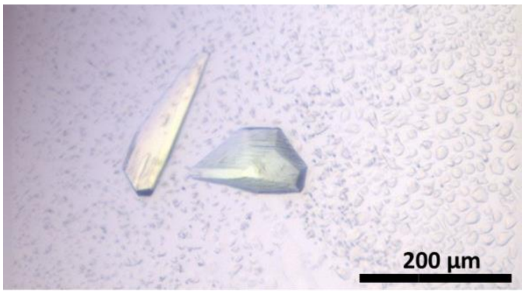


Figure 54: Crystals of PA3127-CTD in condition A1 of the MIDAS screen.

A PA3127-CTD crystal was sent to the BESSY II Synchrotron for data collection on beamline 14.1. 3600 images of 0.1° rotation were collected on a PILATUS 6M detector at a wavelength of λ = 0.9184 Å (Table 38).

Table 38: Data collection statistics for *P. aeruginosa* PA3127-CTD. Values in parentheses are for the highest resolution shell. AcCoA = Acetyl-Coenzyme A.

Dataset	PA3127-CTD + AcCoA
Wavelength (Å) / beamline	0.9184 / BESSY, BL14.1
Resolution range (Å)	39.53-1.70 (1.73-1.7)
Space group	$P3_1$
Unit cell parameters (Å)	78.1 78.1 48.7
(°)	90 90 120
Mosaicity (°)†	0.134
Total No. of measured reflections	376739 (20945)
Unique reflections	36549 (1994)
Multiplicity	10.3 (10.5)
Mean I/σ(I)	17.4 (2.1)
Completeness (%)	100 (100)
R_{meas} (%)	9.1 (128.7)
$R_{p.i.m.}$ (%)	2.8 (39.6)

†Mosaicity values reported by *XDS* (Kabsch, 2010)

The PA3127-CTD crystal was indexed in space group $P3_1$ with cell parameters a = 78.1, b = 78.1, c = 48.7 and γ = 120. Hence, a new cystal form with better diffraction quality has been identified compared to the first crystal form (Popp, 2013).

6.6.6 Structure Determination and Refinement of PA3127-CTD

The structure of PA3127-CTD was successfully determined using *BALBES* (Long *et al.*, 2007) using the structure of the aminoglycoside N-acetyltransferase AAC(6')-Iy in complex with CoA and ribostamycin (PDB entry 1S3Z, Vetting *et al.*, 2004) as MR model. This protein shares 17% sequence identity with PA3127-CTD. After structure determination and initial refinement, the presence of one AcCoA molecule per protein molecule was obvious by positive $m|F_{obs}|-D|F_{calc}|$ density in the active site of PA3127-CTD. The structure of the PA3127-CTD AcCoA complex was finally refined to an R_{work} = 16% and an R_{free} = 19.1% with a *MolProbity* score of 1.15. In the PA3127-CTD structure, 99.3% of the residues are in Ramachandran-favored regions (Table 39).

Table 39: Refinement statistics for *P. aeruginosa* PA3127-CTD in complex with AcCoA. Values in parentheses are for the highest resolution shell.

Dataset	PA3127-CTD + AcCoA
Resolution range (Å)	39.53-1.70 (1.73-1.7)
R_{work} (%)	16.0 (23.8)
R_{free} (%)	19.1 (28.5)
No. of non-H atoms	
Protein	2197
Ion	-
Ligand	102
Water	216
R.m.s. deviations	
Bonds (Å)	0.019
Angles (°)	2.031
Average B factors (Å^2)	
Protein	29
Ion	-
Ligand	44
Water	26
Ramachandran plot	
Favored regions (%)	99.3
Outliers (%)	0.0
MolProbity score#	1.15
PDB entry code	-

#As reported by *MolProbity* (Chen *et al.*, 2010)

6.6.7 Overall Topology of PA3127-CTD

PA3127-CTD shares about 17-20% sequence identity with aminoglycoside N-acetyltransferases and diamine N-acetyltransferases and folds into a typical N-acetyltransferase (Figure 55A). The protein forms an α-β-α sandwich with the central β-sheet consisting of strands β1 to β5. This sheet is flanked by two helices α1 and α2 as well as α3 and α4 on each site.

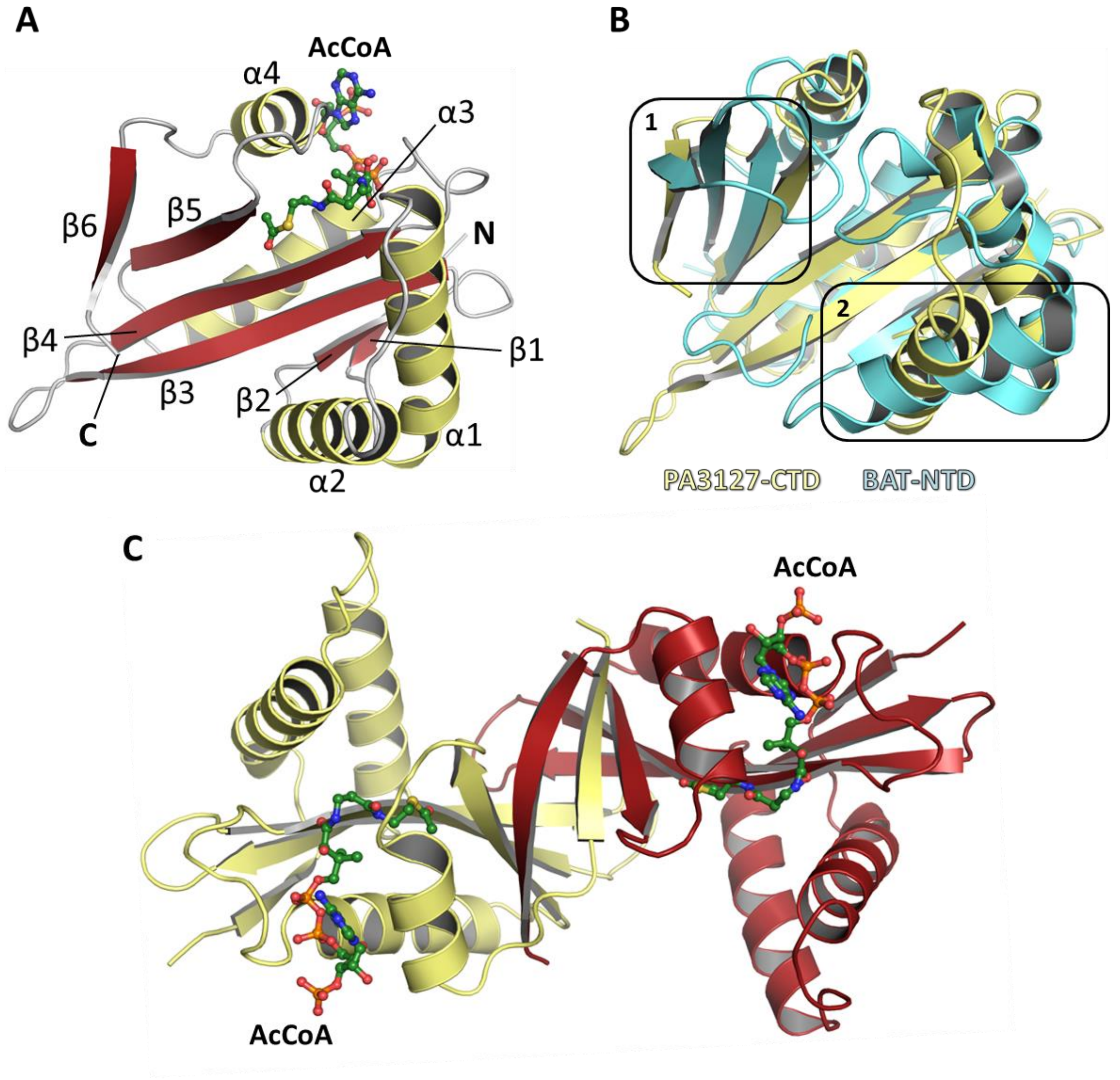


Figure 55: Overall topology of PA3127-CTD in complex with AcCoA. A: The PA3127-CTD monomer folds into a typical N-acetyltransferase domain with four α-helices and six β-strands. B: Superposition of acetyltransferases PA3127-CTD and BAT-NTD. Differences in the three-dimensional structure can be observed in the region of α2 (box 2) and the β5/β6 region (box 1). C: The homodimer of PA3127-CTD is formed by domain swapping between β5 and β6 which provide key interactions in dimer formation.

Interestingly, the three-dimensional structure of PA3127-CTD partially differs from the structure of the N-acetyltransferase domain of BAT from *S. verticillus* (Figure 55B). The arrangement of the C-terminal β5 and β6 strands is different in the proteins (Figure 55B, box 1). Similarly, the arrangement of α1 and α2 is different in the *Streptomyces* protein (Figure 55B, box 2).

The differences in arrangement of the C-terminal β5 and β6 strands of BAT and PA3127 are caused by the domain swapping interactions between β5 and β6 strands from two PA3127-CTD molecules. These interactions lead to dimerization of two PA3127-CTD molecules (Figure 55C). The β-strands tightly interact by hydrogen bonds, mainly between backbone amides and backbone carbonyl groups.

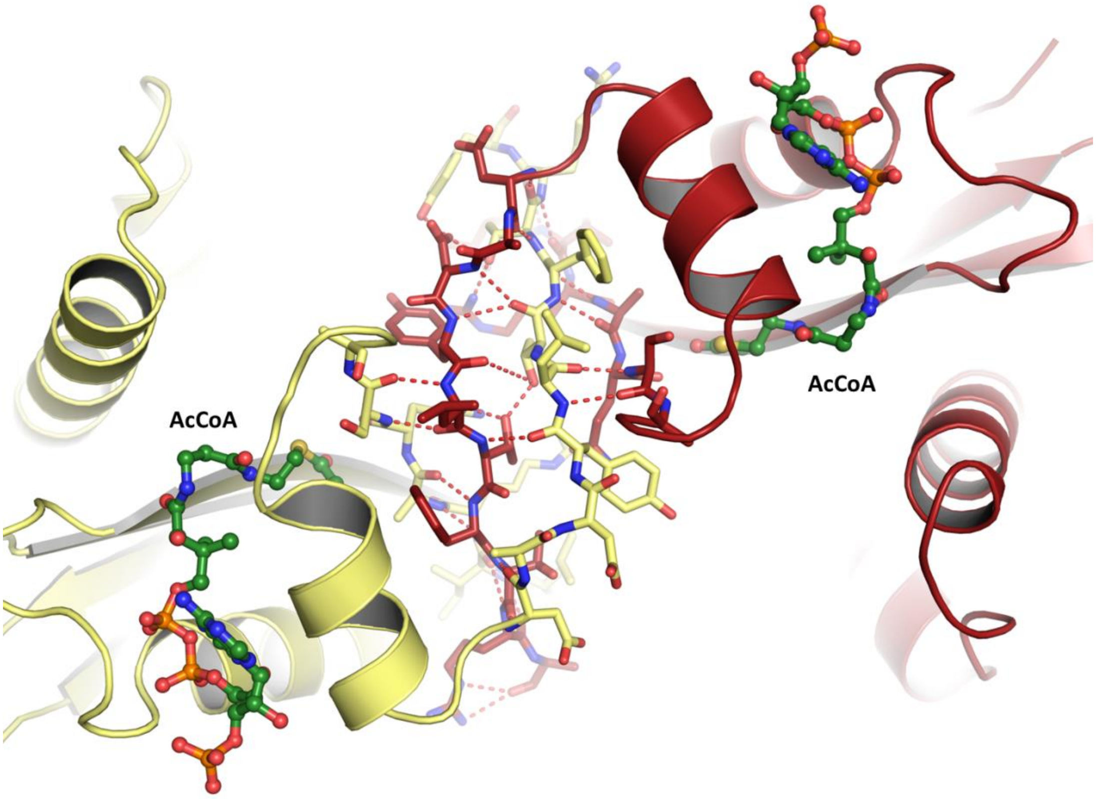


Figure 56: Tight domain swapping interactions by hydrogen bonding between the β5 and β6 strands of two PA3127-CTD molecules. These interacions cause stable dimerization of PA3127-CTD.

6.6.8 AcCoA Binding to PA3127-CTD

PA3127-CTD was cocrystallized with AcCoA which is present in the crystal structure as indicated by positive $m|F_{obs}|-D|F_{calc}|$ electron density (Figure 57A). The cosubstrate is mainly bound by hydrophobic interactions, for instance with L250 or F251, but also stabilized by hydrogen bonds with PA3127-CTD and water molecules. A *CDD* search (Marchler-Bauer *et al.*, 2015) predicts the AcCoA binding site of PA3127-CTD to include residues L206, Y207 and V208 as well as A218, R219 and E220. In fact, AcCoA establishes hydrogen bonds to the backbone carbonyl of L206 and the backbone amides of V208, A218 and R219 as well as to the side chain of R214 (Figure 57B). A full plot of the AcCoA binding interactions is given in Figure 57C.

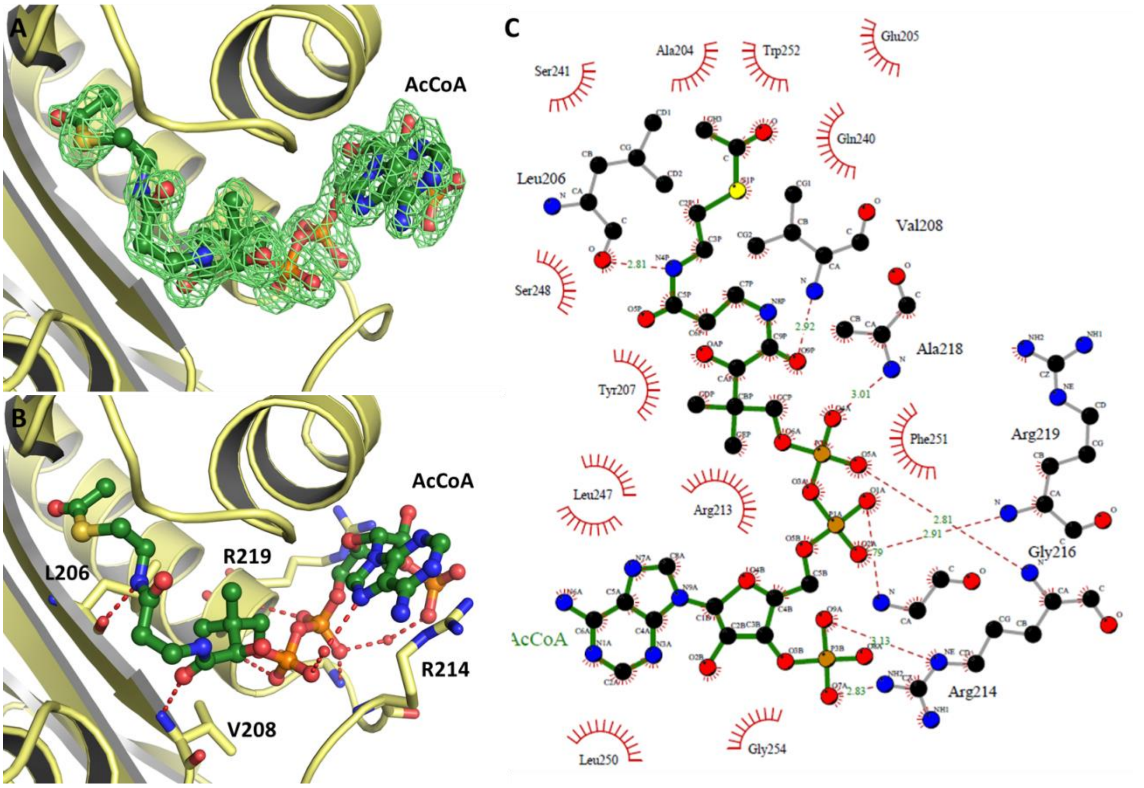
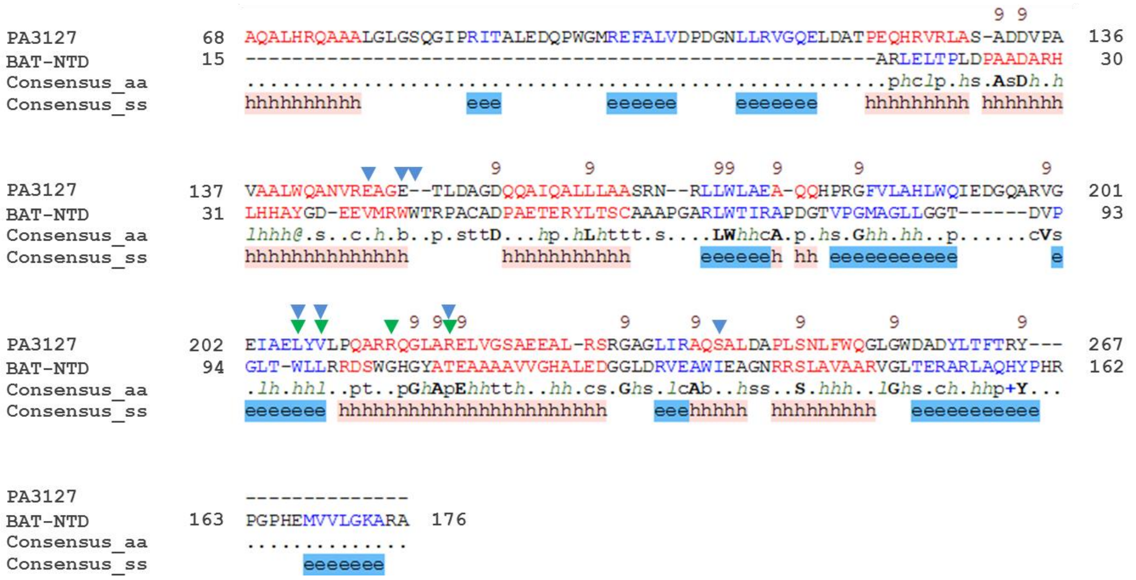


Figure 57: Interactions of PA3127-CTD with AcCoA. A: Positive m|F_{obs}|-D|F_{calc}|density (σ-level = 2) confirms the presence of AcCoA in the PA3127-CTD structure. B: Hydrogen bonding network of AcCoA and PA3127-CTD. Main interactions with the protein are established via the backbone carbonyl of L206, the backbone amide of V208 and the R214 main and side chain amines. C: Plot of the interactions between PA3127-CTD and AcCoA generated using LigPlot+. AcCoA is stabilized by hydrogen bonds and hydrophobic interactions.

6.6.9 Comparison of AcCoA Binding to PA3127 and BAT

The three-dimensional structures of the N-acetyltransferase domains of PA3127 and BAT were submitted to the *PROMALS3D* alignment server (Pei *et al.*, 2008) to identify conserved residues. However, very few and only structurally important residues, are conserved in these domains (Figure 58).

In the AcCoA binding sites, the spatial positions of AcCoA binding site and residues are conserved, however, the residues themselves are not (Figure 58). While AcCoA is mainly stabilized by tryptophans in BAT, PA3127 employes hydrophobic amino acids like valine and leucine as well as polar arginines as AcCoA binding residues. Also, the AcCoA binding mode is different regarding the position of the adenosine moiety of AcCoA (Figure 59A).



```
                                                                                    9 9
PA3127           68 AQALHRQAAALGLGSQGIPRITALEDQPWGMREFALVDPDGNLLRVGQELDATPEQHRVRLAS-ADDVPA 136
BAT-NTD          15 ------------------------------------------------------ARLELTPLDPAADARH  30
Consensus_aa        ......................................................phclp.hs.AsDh.h
Consensus_ss        hhhhhhhhhh         eee         eeeeee     eeeeeee    hhhhhhhhh hhhhhhh

                              ▼  ▼▼
                                         9       9          99   9      9               9
PA3127          137 VAALWQANVREAGE--TLDAGDQQAIQALLLAASRN--RLLWLAEA-QQHPRGFVLAHLWQIEDGQARVG 201
BAT-NTD          31 LHHAYGD-EEVMRWWTRPACADPAETERYLTSCAAAPGARLWTIRAPDGTVPGMAGLLGGT------DVP  93
Consensus_aa        lhhh@.s..c.h.b..p.sttD...hp.hLhttt.s....LWhhcA.p.hs.Ghh.hh..p......cVs
Consensus_ss        hhhhhhhhhhhhhh        hhhhhhhhhhh      eeeeeeh hh eeeeeeeeeee        e

                        ▼ ▼          ▼
                        ▼ ▼     ▼ 9 9▼9             9     9 ▼      9       9          9
PA3127          202 EIAELYVLPQARRQGLARELVGSAEEAL-RSRGAGLIRAQSALDAPLSNLFWQGLGWDADYLTFTRY--- 267
BAT-NTD          94 GLT-WLLRRDSWGHGYATEAAAAVVGHALEDGGLDRVEAWIEAGNRRSLAVAARVGLTERARLAQHYPHR 162
Consensus_aa        .lh.hhl..pt..pGhApEhhtth..hh.cs.Ghs.lcAb..hss..S.hhh..lGhs.ch.hhp+Y...
Consensus_ss        eeeeeee hhhhhhhhhhhhhhhhhhhhhhh    eeehhhh   hhhhhhhhh   eeeeeeeeeee

PA3127              --------------
BAT-NTD         163 PGPHEMVVLGKARA  176
Consensus_aa        ..............
Consensus_ss             eeeeeee
```

Figure 58: Structural alignment of the N-acetyltransferase domains of PA3127 (residues 123-267) and of BAT (residues 9-176). AcCoA binding residues of BAT are marked with a blue triangle, AcCoA binding residues of PA3127 are marked with a green triangle. Structurally conserved residues are marked with a 9.

Few amino acids are conserved in BAT and PA3127 (Figure 58, 9, Figure 59B). These residues are located far away from the AcCoA binding site and propably structurally stabilize the protein. These conserved residues point to a common ancestor of BAT and PA3127, even if the types of residues in their active sites are distinct from each other.

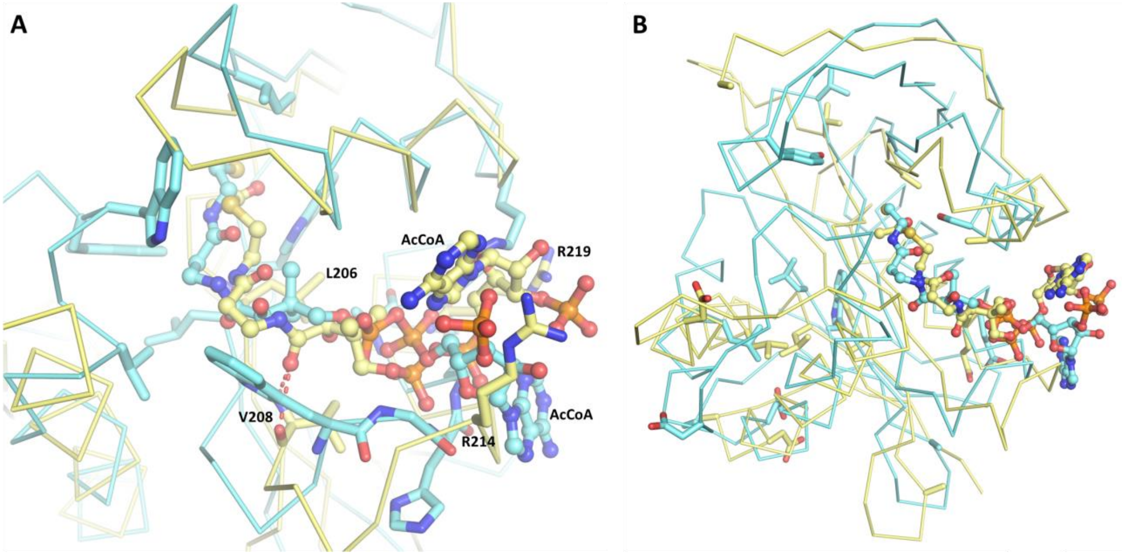


Figure 59: Comparison the N-acetyltransferase domains of BAT (cyan) and PA3127 (yellow). A: AcCoA binding residues. B: Structurally conserved residues in PA3127 and BAT identified by *PROMALS3D* (Pei *et al.*, 2008).

6.6.10 Structure Determination and Refinement of PA3127 Using Both PA3127-NTD and PA3127-CTD as MR Models

The crystal structure of PA3127 could finally be determined by using both the PA3127-NTD and PA3127CTD as MR models. The partial solution obtained using PA3127-NTD before was fixed in *phenix.phaser* and eight copies of a truncated structure of PA3127-CTD (PA3127-CTD lacking strand β6 for the standard cell and PA3127-CTD lacking helix α2 and strand β6 for the dehydrated cell) were searched in the remaining asymmetric unit, which could finally be located by *phenix.phaser*.

Table 40: Refinement statistics for *P. aeruginosa* PA3127 in complex with residual AcCoA/CoA. Values in parentheses are for the highest resolution shell.

Dataset	PA3127 – standard cell	PA3127 – dehydrated cell
Resolution range (Å)	96.58-2.95 (3.04-2.95)	49.03-2.48 (2.64-2.48)
R_{work} (%)	23.7 (32.9)	30.2 (36.4)
R_{free} (%)	29.1 (37.8)	35.3 (44.0)
No. of non-H atoms		
Protein	15471	15245
Ion	10	-
Ligand	312	-
Water	100	58
R.m.s. deviations		
Bonds (Å)	0.003	0.006
Angles (°)	0.867	1.149
Average B factors (Å^2)		
Protein	77	83
Ion	44	-
Ligand	73	-
Water	36	72
Ramachandran plot		
Favored regions (%)	93.1	85.7
Outliers (%)	1.1	5.2
MolProbity score#	2.05	2.9
PDB entry code	-	-

#As reported by *MolProbity* (Chen *et al.*, 2010)

Interestingly, AcCoA/CoA was found in the standard cell crystal structure of PA3127. As neither the cosubstrate nor the product was added to the crystallization experiment, it must originate from the protein production organism *E. coli* and must have been bound to the protein during the whole purification process.

The structures of PA3127 were finally refined to an R_{work} = 23.7% and an R_{free} = 29.1% with a *MolProbity* score of 2.05 for the standard cell, and an R_{work} = 30.2% and an R_{free} = 35.3% with a *MolProbity* score of 2.9 for the dehydrated cell (Table 40). In the crystals of PA3127 with the standard and the dehydrated cell, 93.1% and 85.7% of the residues are found in Ramachandran-favored regions, respectively. Possibly, crystal dehydration and shrinking of the unit cell favor unusual rotamers and Ramachandran outliers.

6.6.11 Overall Topology of PA3127

Crystal packing of the PA3127 crystals with eight molecules in the asymmetric unit is given in Figure 60.

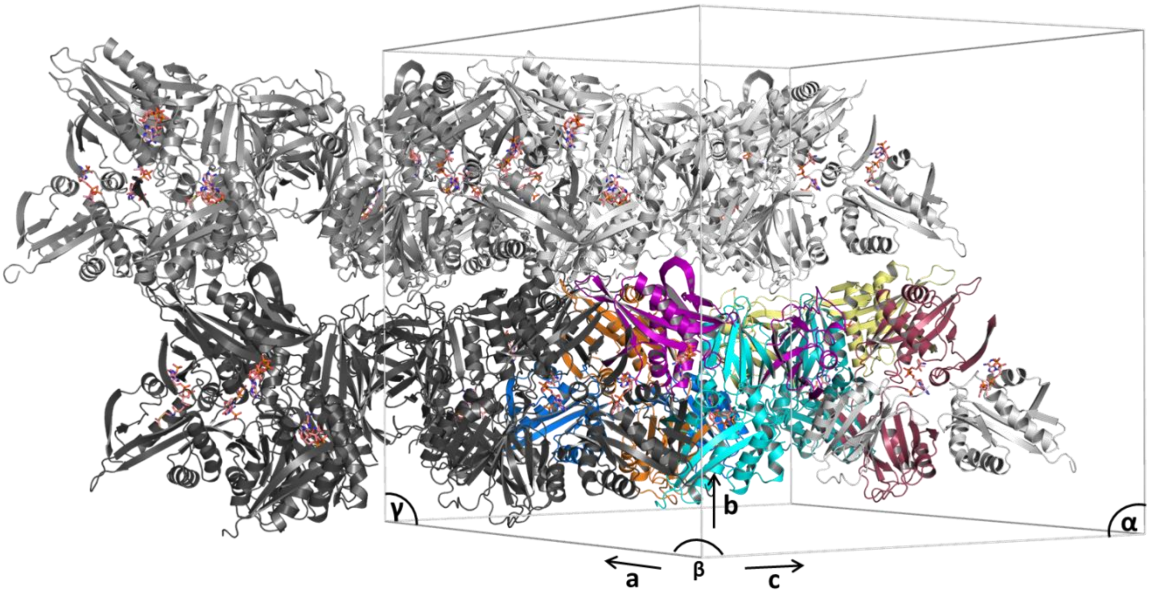


Figure 60: Crystal packing of the PA3127 crystals with one asymmetric unit containing eight PA3127 molecules as colored cartoon representation, and other three asymmetric units in grey. The unit cell with parameters a = 106.9 Å, b = 121 Å, c = 116.2 Å and α = 90°, β = 115.35° and γ = 90° is shown in green.

In the crystals, the PA3127 molecules are arranged with the βαβββ-module domains PA3127-NTD pointing towards the core of the asymmetric unit, and the eight N-acetyltransferase domains PA3127-CTD pointing towards the outer boundaries (Figure 61). As discussed previously, the N-terminal domain of PA3127 folds into the typical two-βαβββ-module architecture and dimerizes via edge-to-edge contacts between the β1 and β5 strands of two domains to most probably form the binding site for the bis-thiazole moiety of bleomycin. The βαβββ-module domain dimers are found within one asymmetric unit (Figure 61, e. g. cyan/violet). Similarly, the C-terminal N-acetyltransferase domain of PA3127 folds into a typical structure of the N-acetyltransferase superfamily. This domain dimerizes via domain swapping interactions between the β5 and β6 strands of two N-acetyltransferase domains of PA3127. The N-acetyltransferase domain dimers are formed by two domains for different asymmetric units (Figure 61, e. g. blue/green).

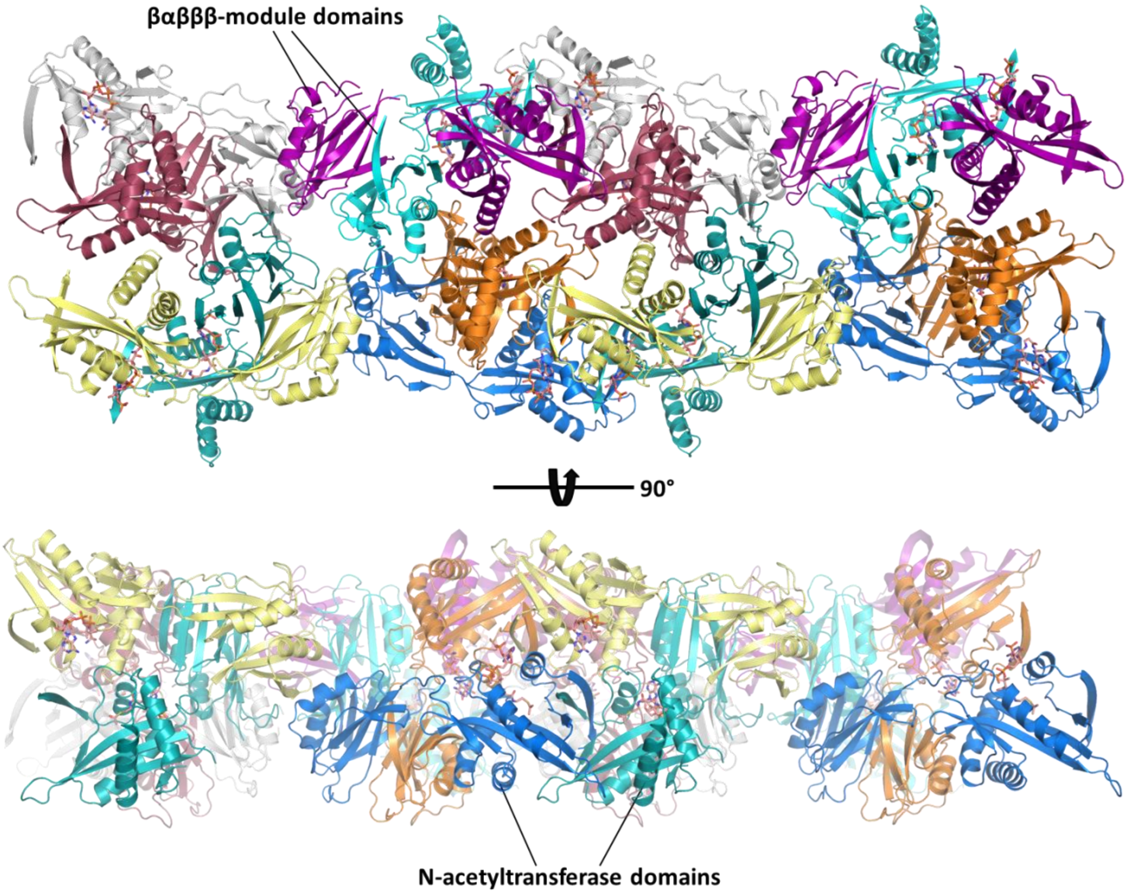


Figure 61: Two asymmetric units of the PA3127 crystals, each containing eight molecules (colored differently), demonstrate that important contacts between single PA3127 molecules are established by dimerization interactions between the βαβββ-module domains within one asymmetric unit, and by the N-acetyltransferase domains, between molecules from two distinct asymmetric units. AcCoA bound to the N-acetyltransferase domain is shown as stick representation.

If PA3127 is assumed to be a homodimer as all typical βαβββ-module proteins with two modules per molecule and also BAT from *S. verticillus*, the latter can be compared to

PA3127 and similarities as well as differences between these bleomycin-N-acetyltransferases can be pointed out (Figure 62). In BAT from *S. verticillus*, the two protein molecules of the dimer are symmetrically arranged and can be mapped to each other by superposition of the molecules.

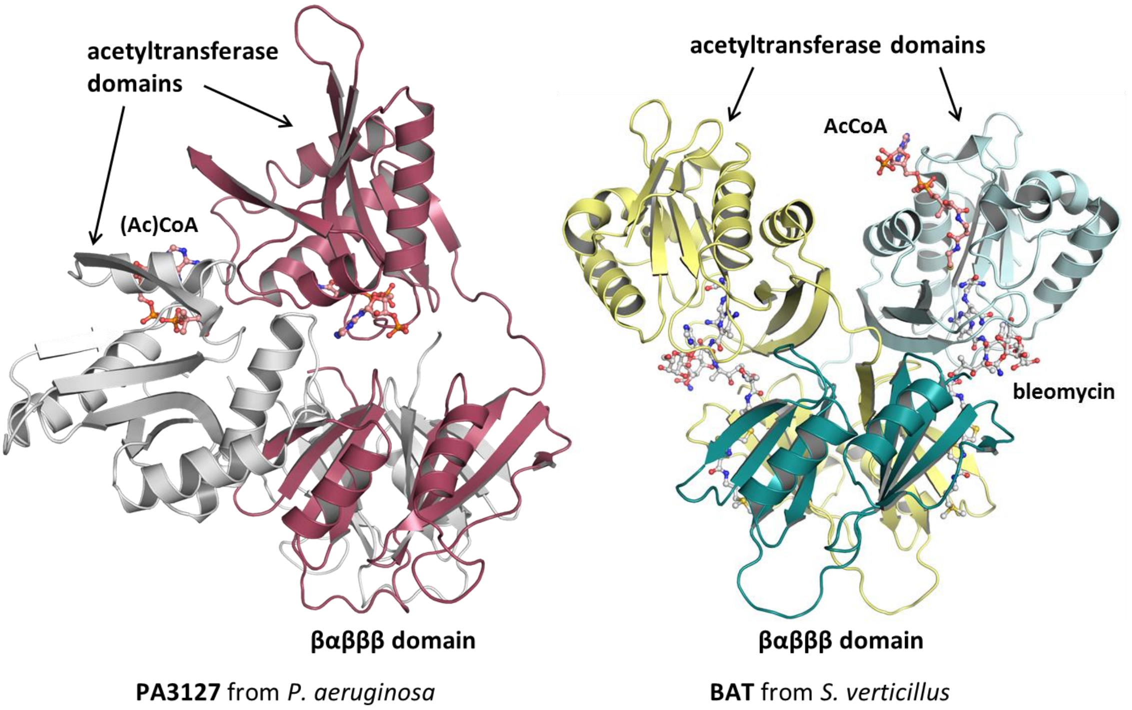


Figure 62: Comparison between a PA3127 dimer from *P. aeruginosa* and a BAT dimer from *S. verticillus* (PDB entry 2ZW7, Oda *et al.*, 2010). While the BAT dimer is symmetrical and the domains are arranged identically in the monomers, the PA3127 dimer is not symmetrical. One N-acetyltransferase domain is located in line with its βαβββ-module domain (red). The second N-acetyltransferase is kinked in respect to its βαβββ-module domain (grey).

In contrast, the two molecules of the PA3127 dimer are not symmetrical and cannot be superimposed to each other. One βαβββ-module domain is in line with its corresponding N-acetyltransferase domain (Figure 62, red), while the second N-acetyltransferase domain is kinked in respect to its corresponding βαβββ-module domain (Figure 62, grey). This asymmetric arrangement is repeated in the additional three dimers of the asymmetric unit, which means that every dimer consists of a stretched and a kinked PA3127 molecule, thereby causing in total four PA3127 molecules to be stretched and the other four PA3127 molecules to be kinked. This observation explains why the number of molecules in the PA3127 asymmetric unit could not be determined by calculating a self-rotation function (Winn *et al.*, 2011). Additionally, it explains why the PA3127 structure could not be phased by applying the NCS operators of the βαβββ-module domains to the whole protein, thereby improving the electron densities and facilitating structure determination and refinement.

However, the orientations of the PA3127 domains relative to each other might be an artefact of crystal packing. In order to investigate potential overall conformation and the oligomeric state of PA3127 in solution, the protein was subjected to SAXS experiments.

6.6.12 SAXS Reveals PA3127 to be a Tetramer in Solution

Dimerization of PA3127 can occur via the βαβββ-module domain PA3127-NTD as well as the N-acetyltransferase domain PA33127-CTD. In agreement with these observations, previous analysis of PA3127 indicated a higher oligomerization state in size exclusion chromatography and in blue native PAGE (Bittner, 2012). For these reasons, PA3127 was subjected to SAXS experiments on BM29 at the ESRF (Grenoble, France). Solution scattering patterns of a PA3127 dilution series (15, 5, 2.5, 1.25, 0.75 and 0.375 mg/mL in SEC buffer) were collected at a wavelength of λ = 0.9919 Å. In case of PA3127, the best sample quality and monodispersity (91.8%) was achieved at a protein concentration of 1.25 mg/mL.

In order to define the oligomeric state of PA3127, its crystal structure was submitted to the *PISA* server (Krissinel & Henrick, 2007) to determine all possible stable oligomeric assemblies. Theoretical scattering curves of these assemblies (dimer, tetramer and octamer) were calculated and compared to the experimentally determined scattering curve of PA3127 (Figure 63). The comparison of the calculated and the experimentally determined SAXS curves yielded curve fits with different fitting values of X = 16.215, X = 21.971, X = 5.273 and X = 5.509 for the octamer, dimer and both tetramers (which are structurally identical), respectively. Hence, PA3127 is very likely to be a tetramer in solution. The calculated tetramer possesses an overall surface area of 47120 Å^2 with a buried surface of 1790 Å^2 according to *PISA* (Krissinel & Henrick, 2007).

The experimentally determined SAXS curve and the tetramer of PA3127 were then used to calculate the SAXS envelope of the protein which is given in Figure 64.

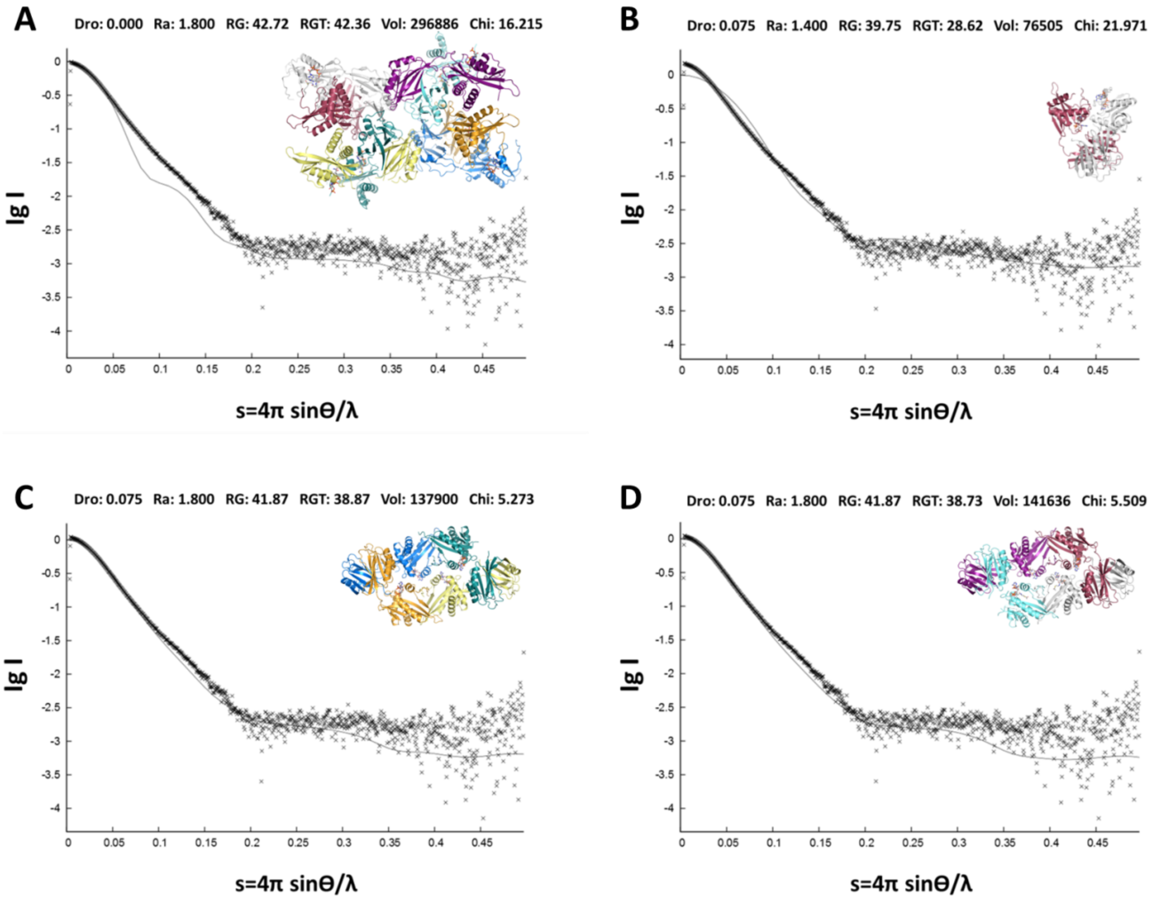


Figure 63: SAXS analysis of the oligomerization state of PA3127. Theoretical scattering curves of the PA3127 octamer (A), dimer (B) and tetramer (C: chains BCEF; D: chains ADGH, structurally identical) were compared to the experimentally determined SAXS curve. The best fit is achieved using a PA3127 tetramer, resulting in a Χ-value of 5.273 (C).

The PA3127 tetramer is formed by a dimer of dimers. One dimerization occurs via two βαβββ-module domains (Figure 64, blue and orange). The N-acetyltransferase belonging to these two βαβββ-module domains then each dimerize with one N-acetyltransferase domain from a second PA3127 dimer (Figure 64, yellow and green) to form the full tetramer.

The PA3127 tetramer fits the computed rhombic-shaped SAXS envelope. However, the βαβββ-module domains slightly stick out of this envelope, causing a fitting value of Χ = 5.273. Probably, the conformation of the tetramer is flat in solution, like the SAXS envelope, but bent in the crystal structure. The βαβββ-module domains probably rotate into the SAXS envelope. This rotation is possible because of the ten amino acid long linkers between the βαβββ-module and the N-acetyltransferase domains.

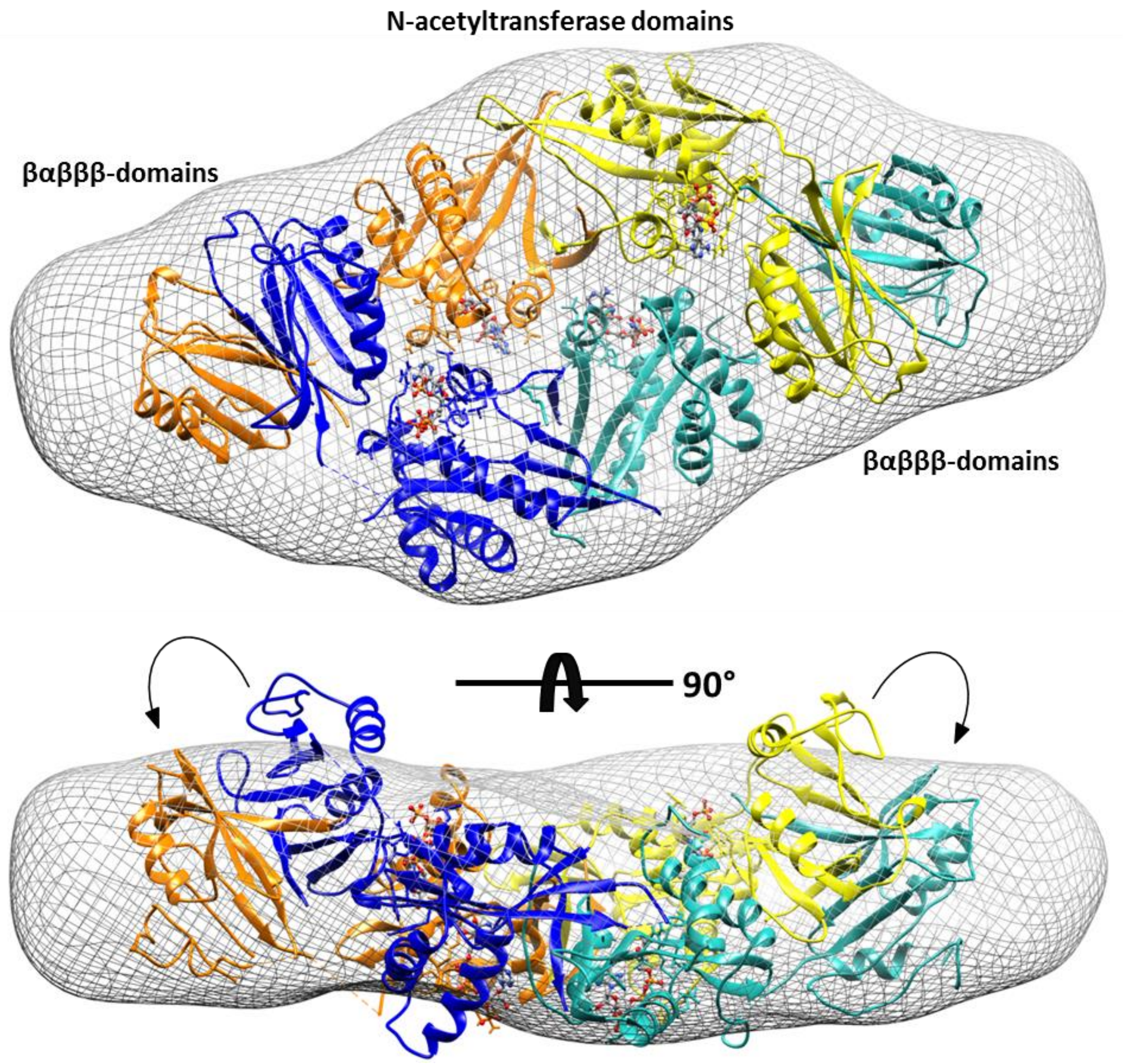


Figure 64: SAXS envelope of the PA3127 tetramer (in complex with AcCoA) which is a dimer of dimers. Its formation occurs by dimerization of two βαβββ-module domain, whose N-acetyltransferase domains dimerize with two N-acetyltransferase domains from another PA3127 dimer. The βαβββ-module domains of the tetramer are slightly bent out of the SAXS envelope, which is probably caused by crystal packing. In solution, the tetramer might relax and form a plane quaternary assembly.

It is not clear if the tetramer is essential for the physiological function of PA3127. Possibly it is not essential for the catalytic activity, which might take place within a dimer (for instance Figure 64, blue and orange). A similar arrangement was observed in the tryptophanase from *E. coli* (Kogan *et al.*, 2009). Hence, the tetramer most probably represents the true PA3127 oligomer in solution, even if it is not catalytically essential.

6.6.13 Spot Plate Assay with PA3127 and Its Operon

Spot plate assays using both *E. coli* and *P. aeruginosa* were performed in order to investigate the potential role of PA3127, PA3128 and both PA3127 and PA3128 (as whole operon) to mediate bleomycin resistance *in vivo*. For this purpose, *P. aeruginosa* PA14 wild type and three ΔPA3127 transposon mutants were spotted on bleomycin-doted agar plates. Similarly, *P. aeruginosa* PAO1 were transformed with pHERD30T plasmids encoding either no protein (as a control) or PA3127, and spotted on bleomycin-agar together with *P. aeruginosa* PAO1 wild type. However, no differences in growth could be observed in the experiments with *P. aeruginosa* PAO1 and PA14. This might be due to unknown factors such as efflux pumps that might act in concert with PA3127.

For this reason, *E. coli* was used for the bleomycin spot plate experiments, as this organism does not possess a bleomycin resistance protein. *E. coli* carrying pET19m plasmids encoding either no protein (as a control), PA3127, PA3128 or both PA3127-PA3128 (encoded as the *P. aeruginosa* operon) were spotted on bleomycin-doted M9 agar plates (Figure 65).

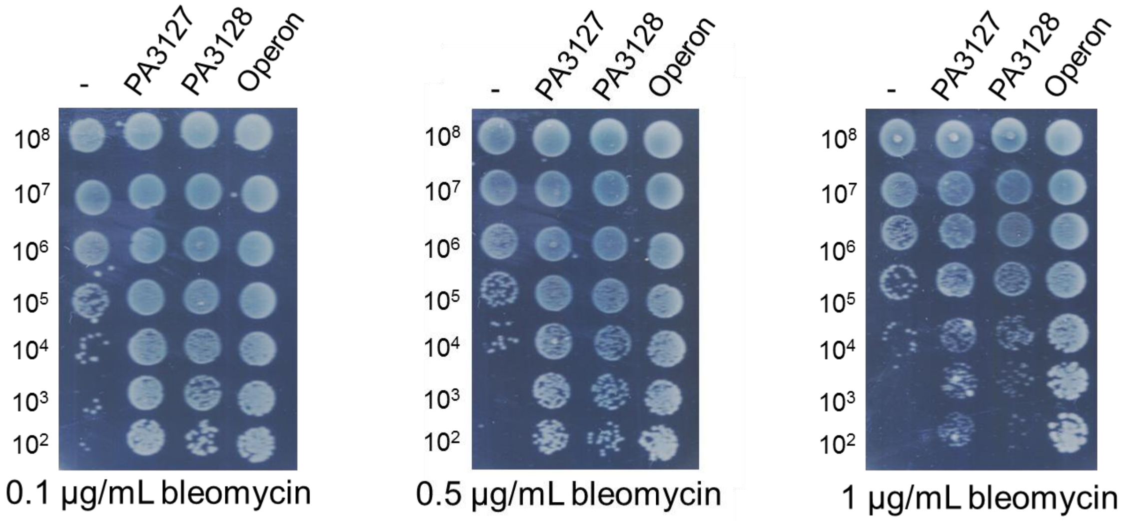


Figure 65: Spot plate assays with bleomycin using *E. coli* (indicated as number of spotted cells/mL culture*)*. Three different concentrations of bleomycin (0.1 µg/mL, 0.5 µg/mL and 1 µg/mL) were used to test the ability of PA3127 to render *E. coli* resistant against bleomycin. The most significant effect is visible at a concentration of 1 µg/mL bleomycin. PA3127 alone makes *E.coli* resistant against bleomycin resistance, while the entire operon PA3127-PA3128 makes *E.coli* even more resistant against this antibiotic.

Three different concentrations of bleomycin (0.1 µg/mL, 0.5 µg/mL and 1 µg/mL) were used to test the ability of PA3127 to render *E. coli* resistant against bleomycin. The most significant effect is visible at the highest concentration of bleomycin (1 µg/mL, Figure 65). *E. coli* carrying an empty pET19m vector only grows to a dilution of 10^4 cells/mL, while *E. coli* carrying a vector pET19m_PA3127 or pET19m_PA3128 grow to a dilution of

10^2 cells/mL. However, more colonies of *E.coli* producing PA3127, the proposed bleomycin N-acetyltransferase, grow on bleomycin-doted agar than of *E. coli* producing PA3128, a proposed oxidoreductase encoded in an operon with PA3127. The best growth was achieved with *E. coli* carrying a plasmid pET19m_PA3127-PA3128, producing both proteins of the operon. These data demonstrate that PA3127 is a bleomycin resistance protein that is able to render *E. coli* resistant against this antibiotic. Furthermore, the *P. aeruginosa* operon *PA3127-PA3128* is a bleomycin resistance determinant. Together with the biophysical and structural data obtained in this study, PA3127 is assigned as a bleomycin-N-acetyltransferase. The function of PA3128 has to be identified in future studies.

6.6.14 Bleomycin Binding

As no crystal structure of PA3127 with bleomycin could be obtained, the binding mode of bleomycin to PA3127 remains elusive. As the mode of bleomycin binding is known in BAT from *S. verticillus*, it may be used to derive possible bleomycin binding modes for PA3127. In BAT, bleomycin spans a distance of 32 Å in its bound conformation *(Oda et al.*, 2010), while an elongated bleomycin may be able to span a distance of even 40 Å.

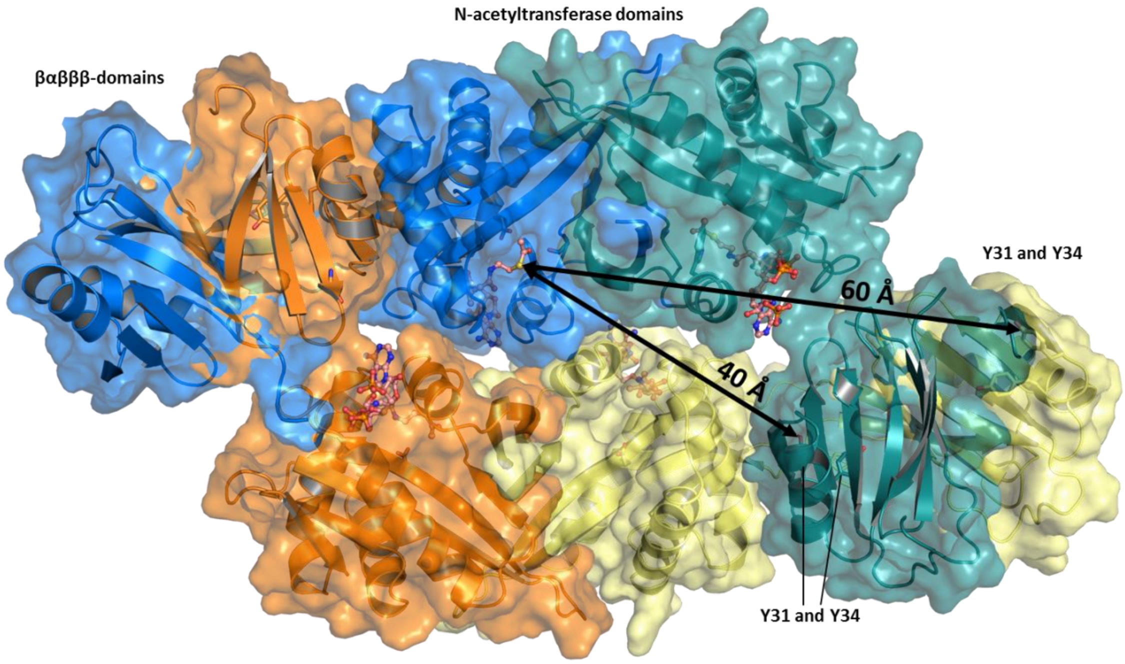


Figure 66: Evaluation of potential bleomycin binding to the PA3127 tetramer. Direct distances between acetyl groups and Y31 and Y34 of different dimers of the PA3127 tetramer were measured in *PyMOL* (DeLano, 2002). These distances are too long to be bridged by bleomycin. Also, the real distances would be even longer, as the direct distances cross through the protein.

In order to determine the possible binding conformations of bleomycin in PA3127, the distances between its βαβββ-module domains and the acetyl groups of bound AcCoA were measured in *PyMOL* (DeLano, 2002). However, distances between acetyl groups of AcCoA molecules bound to one dimer of the tetramer and Y31 or Y34 belonging to the other dimer of the tetramer are too long (40 and 60 Å) to be bridged by bleomycin (Figure 66). Four different distances can be measured within a PA3127 tetramer (Figure 67).

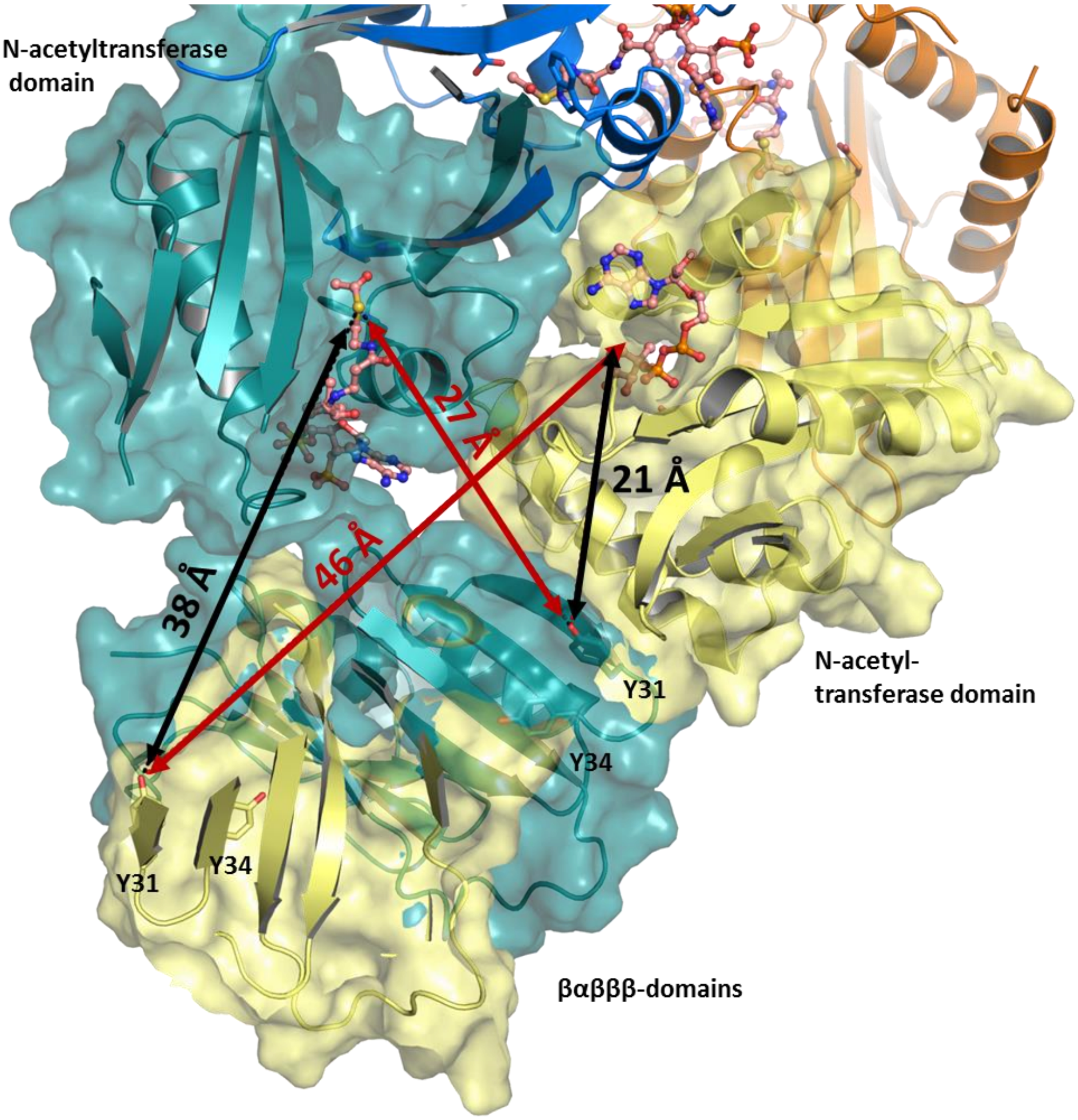


Figure 67: Evaluation of potential bleomycin binding to PA3127 dimer. Direct distances between acetyl groups and Y31 and Y34 of different dimers of the PA3127 tetramer were measured in *PyMOL*. Two of these distances are too long to be bridged by bleomycin, while two are short enough. However, steric hindrance does not allow the bleomycin molecule to bind to Y31, Y34 and the acetyl group of AcCoA within the same monomer (red arrows).

However, the asymmetry in the crystallographic PA3127 tetramer is probably absent in solution, as indicated by SAXS analysis. In solution, the βαβββ-module domains probably undergo a rotation, leading to a plane and rhombic-shaped tetramer. Still, a potential bleomycin binding groove is visible in the PA3127 tetramer (Figure 68). This groove could be shortened by the rotation of the βαβββ-module domains, thereby allowing bleomycin binding in a stretched conformation as observed in BAT from *S. verticillus* (Oda *et al.*, 2010). In this conformation, the primary amine of bleomycin would be able to position itself in the vicinity of the AcCoA acetyl group.

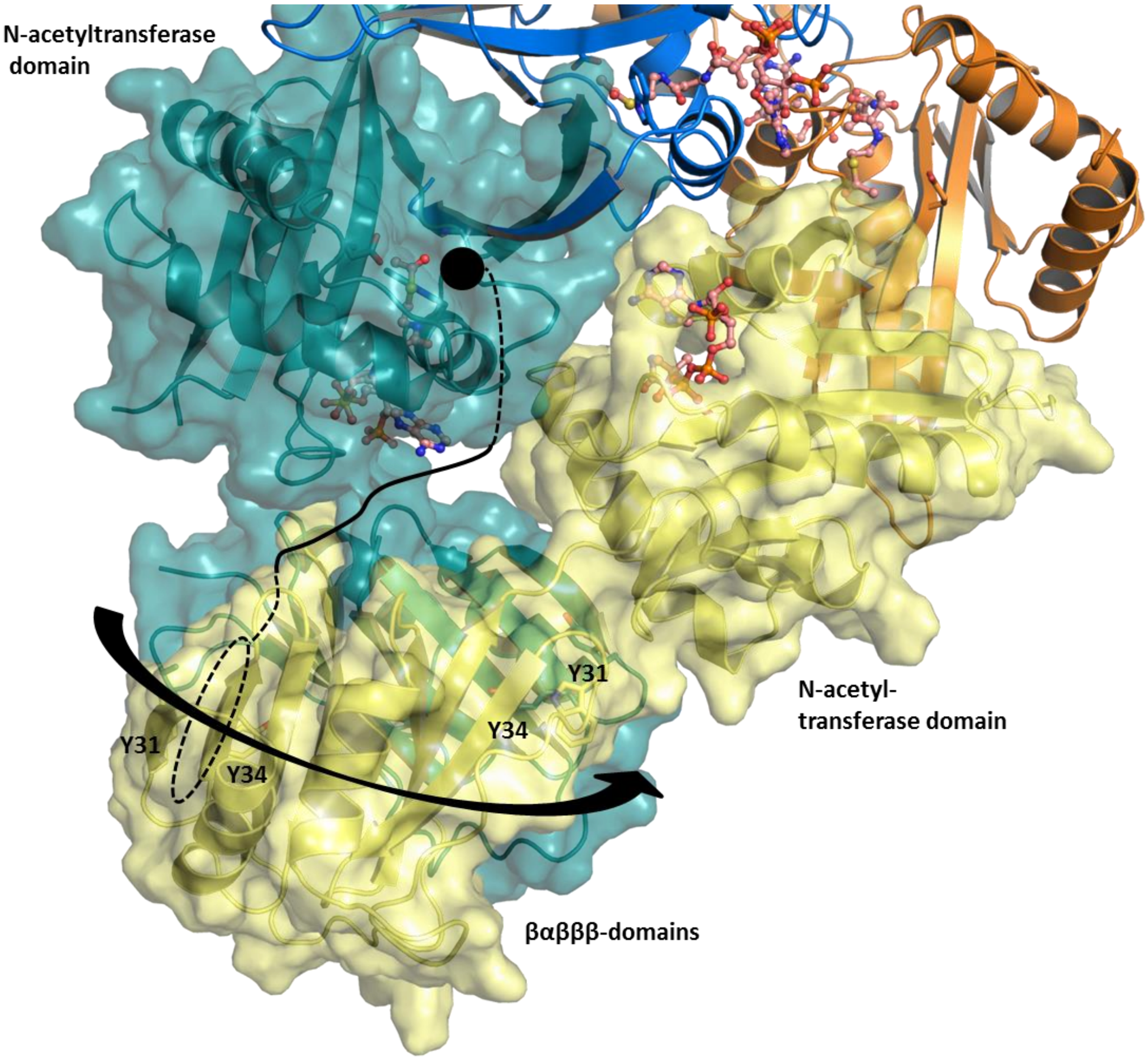


Figure 68: Potential movement of the βαβββ-domains of PA3127 and potential bleomycin binding mode. The βαβββ-domains probably have to turn around to shorten the distance between the bis-thiazole moiety and primary amine binding sites of bleomycin. Then, bleomycin would be able to bind in its stretched conformation.

Interestingly, two distinct, stretched bleomycin binding conformations have been observed in BAT from *S. verticillus* (Figure 62, Oda *et al.*, 2010). The stretched conformations tear the metal ligands of bleomycin, including the primary amine, apart from each other. Hence, BAT only binds the metal-free from of bleomycin (Oda *et al.*, 2010). This scenario is most likely present in PA3127. In contrast, BLMT and BLMA, two other bleomycin-binding proteins, both bind the metal-bound form of bleomycin (Maruyama *et al.*, 2001; Sugiyama *et al.*, 2002).

6.6.15 Mechanism of Acetyl Transfer from AcCoA to Bleomycin

Using the structural data obtained in this study and the BAT structure from *S. verticillus*, it is still difficult to propose a molecular model of the active PA3127-AcCoA-bleomycin complex and a mechanism of acetyl transfer from AcCoA to bleomycin.

The acetylation of bleomycin by PA3127 most probably occurs in an ordered sequential mechanism: first, AcCoA is bound to the protein, then bleomycin is bound, the acetyl transfer occurs, acetyl-bleomycin is released and finally CoA is released. One observation supporting this theory is the presence of AcCoA in the PA3127 crystal structure, although AcCoA was not added to the crystallization experiment. The same ordered mechanism was also proposed for BAT (Oda *et al.*, 2010), but could not be confirmed by kinetic analysis. Possibly, the protein is permanently loaded with AcCoA in order to immediately catalyze acetyl transfer upon bleomycin binding, as the AcCoA concentration is about 170 mM in exponential phase and between 40-80 mM in stationary phase in *E. coli* (Park *et al.*, 2011). These amounts are more than 10-fold higher than the $K_D = 529 \pm 33$ nM of AcCoA and PA3127. Furthermore, the 10-fold higher affinity of AcCoA compared to CoA ensures correct loading of the protein with the substrate. If bleomycin is present, its large binding surface on PA3127 probably ensures immediate binding to the PA3127-AcCoA complex.

As in BAT from *S. verticillus*, the primary amine of bleomycin is probably positioned in close proximity of the acetyl group of AcCoA. This proximity ensures the efficient transfer of the acetyl group to bleomycin (Oda *et al.*, 2010), which occurs by nucleophilic attack. This nucleophilic attack is probably a direct attack of the primary amine of bleomycin to AcCoA, as suggested for other members of the N-acetyltransferase superfamily proteins (De Angelis *et al.*, 1998; Tanner *et al.*, 2000a; b). In order to attack as a nucleophile, the primary amine of bleomycin has to be deprotonated (Oda *et al.*, 2010). Proton abstraction can either occur by an aspartate or glutamate residue in the vicinity of the amine (Trievel *et al.*, 1999; He *et al.*, 2003), or by water molecules that act as a proton wire (Hickman *et al.*, 1999).

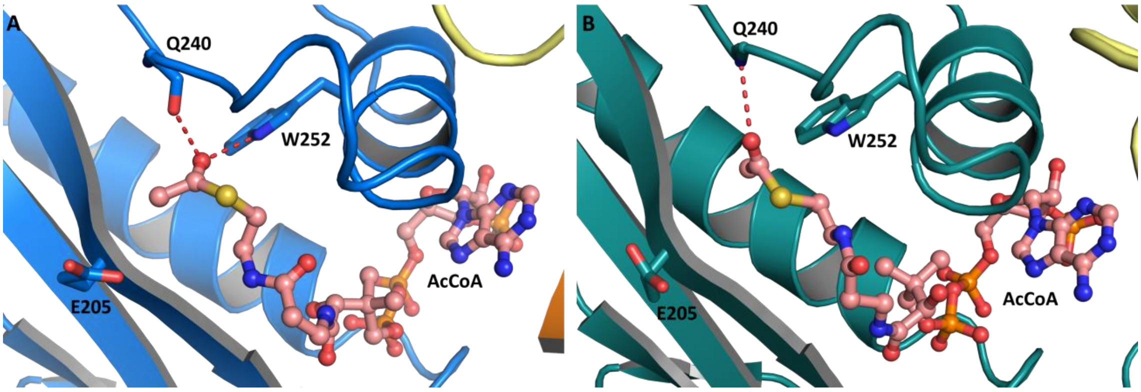


Figure 69: Two AcCoA conformations observed in PA3127. The carbonyl oxygen of AcCoA might be stabilized by the backbone amide or carbonyl of Q240, although the latter is unlikely, and the indole nitrogen of W252. However, the orientation of the acetyl group cannot be determined without uncertainty. The only amino acid that may potentially abstract a proton from the primary amine of bleomycin is E205. However, as the binding mode of bleomycin is elusive, no data to corroborate this hypothesis are available.

In PA3127, however, the only residue that is potentially able to abstract a proton from the primary amine of bleomycin and that is in the vicinity of the acetyl group is E205 (Figure 69). As the binding mode of bleomycin remains elusive, this hypothesis might only be proven by an E205A mutant that might not be possible to abstract the proton anymore and therefore might be inactive. Similarly, in the PA3127 structure, the orientation of the acetyl group of AcCoA cannot be determined with certainty. Possibly, the carbonyl oxygen is stabilized by hydrogen bonds to the indole nitrogen of W252 or the backbone amine or carbonyl of Q40, although the latter is unlikely. In contrast, the acetyl group of AcCoA does not seem to establish any hydrogen bonds in the PA3127-CTD structure.

Taken together, a cocrystal structure of PA3127 and bleomycin would support a detailed analysis of the mechanism of PA3127. Still, PA3127 is definitely a bleomycin-N-acetyl transferase, as demonstrated by the spot plate assays in *E. coli*. In this study, ITC experiments, X-ray crystallography, SAXS and *in vivo* microbiological assays were successfully combined to corroborate this initial hypothesis.

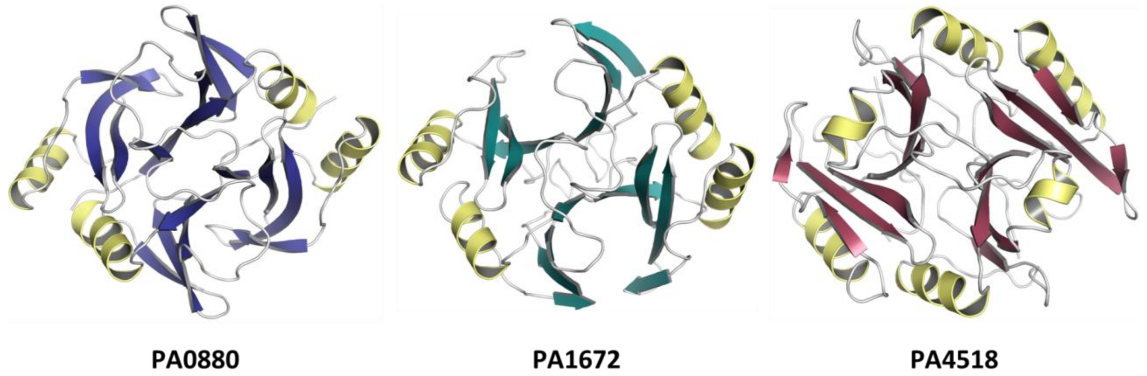
7 Conclusions and Outlook

In this study, 22 βαβββ-module resistance proteins were identified in *P. aeruginosa* PAO1 using a bioinformatic approach, namely a *BackPhyre* search with the crystals structure of EhpR (Yu *et al.*, 2011) in the *P. aeruginosa* genome. The 22 identified βαβββ-module resistance proteins were sorted into three distinct classes using a sequence and structural alignment in *PROMALS3D* (Pei *et al.*, 2008). Class I is constituted by twelve metal-binding βαβββ-module resistance proteins with two conserved histidines and a glutamate residue as the binding motif for divalent metal cations. Class II is formed by four aromatic compound binding proteins, PA0803, PA1353, PA4641 and PA3127. The latter is distinct from the other three because it possesses an additional N-acetyltransferase domain. The aromatic compound binding proteins possess two conserved aromatic residues between which aromatic compounds are able to bind via π-stacking interactions. Class III is constituted by six proteins of unknown function which do neither possess the metal-binding nor the aromatic compound-binding residues.

7.1 βαβββ-Module Resistance Proteins of Unknown Function - PA4183, PA1672, PA4518 and PA0880

The class III βαβββ-module resistance proteins PA0880, PA1672, PA4183 and PA4518 have been cloned and purified with high yield and purity. The proteins were subjected to crystallization, and crystals were indeed obtained of the full proteins PA0880, PA1672 and PA4518 as well as of a truncated version of PA418. However, the latter crystals did not show any diffraction. In contrast, the crystals of PA0880, PA1672 and PA4518 diffracted to better than 2 Å resolution, allowing data collection and structure determination (Figure 70). The proteins are typical βαβββ-module resistance proteins and form homodimers with a binding or active site at each dimer interface.

Functional characterization of these proteins will be a main task in the future. Several bioinformatic approaches using their sequence and structure failed, as they share low sequence homology and very few residues seem to determine their catalytic activity and function.



PA0880 PA1672 PA4518

Figure 70: Crystal structures of the homodimers of the class III βαβββ-module resistance proteins of unknown function PA0880, PA1672 and PA4518.

PA0880 is a good starting point for functional studies, as it is most probably involved in the itaconate degradation pathway (Sasikaran *et al.*, 2014). For PA0880, further potential substrates to be tested in future binding experiments are acrylate, cis-aconitate, methylsuccinate, (*R*)-citramalate as well as (*R*)-3-, (*S*)-3- and 4-hydroxybutyrate. Additionally, CoA derivatives of the substances tested in this study as well as additional CoA derivatives, for instance acetyl-, succinyl-, (3*S*)-malyl-, (3*S*)-citryl-, β-methyl-malyl-, methylcrotonyl-, mesaconyl-, (*S*)-citramalyl- or (*R*,*S*)-3-hydroxy-3-methylglutaryl-CoA could be assayed for affinity to PA0880. However, ITC might not be the optimal method to screen for potential ligands and substrates, as it is time-consuming and requires high amounts of sample, in contrast to a thermal shift assay or microscale thermophoresis (MST) which are probably more suitable for fast screening as they are faster and require less amounts of material.

7.2 Future Studies on Proteins of Unknown Function

In order to functionally characterize the βαβββ-module resistance proteins of unknown function, first, bioinformatic approaches have to be explored, for instance the genomic context of the proteins has to be examined. Furthermore, the function of the proteins may be studied by transcriptomic approaches, which may be able to identify genes that are transcribed together with the gene of interest under certain environmental conditions. Additionally, proteomics may be able to identify protein interaction partners of PA0880, PA1672 and PA4518. Metabolomic approaches will also play an important role in identifying the function of the proteins. This functional identification may either be accomplished by using *P. aeruginosa* transposon or deletion mutants of the gene of interest, and preparing cell-free metabolite extracts that can be analyzed via HPLC-MS. Another method that can be applied to identify small metabolites interfering with proteins of unknown functions is the *BiOLOG* System (BiOLOG, Hayward, CA, U.S.A.). In this system, bacterial phenotypes are screened in a 96-well plate format against different media and nutrients. By doing so, the genotype of a cell can be correlated with a phenotype under certain environmental conditions and the cell's metabolic and chemical sensitivity properties can be characterized.

7.3 Pyocyanin Resistance in *P. aeruginosa*

The class II βαβββ-module resistance proteins PA0803, PA1353 and PA4641 have been structurally and biophysically characterized in studies prior to this work (Yu, 2009; Kalawy-Fansa, 2010). They have been identified as potential pyocyanin resistance determinants in *P. aeruginosa*, and ITC data as well as cocrystal structures of PA1353 and PA4641 with pyocyanin were available (Kalawy-Fansa, 2010). In this study, the cocrystal structure of PA0803 with pyocyanin was determined (Figure 71).

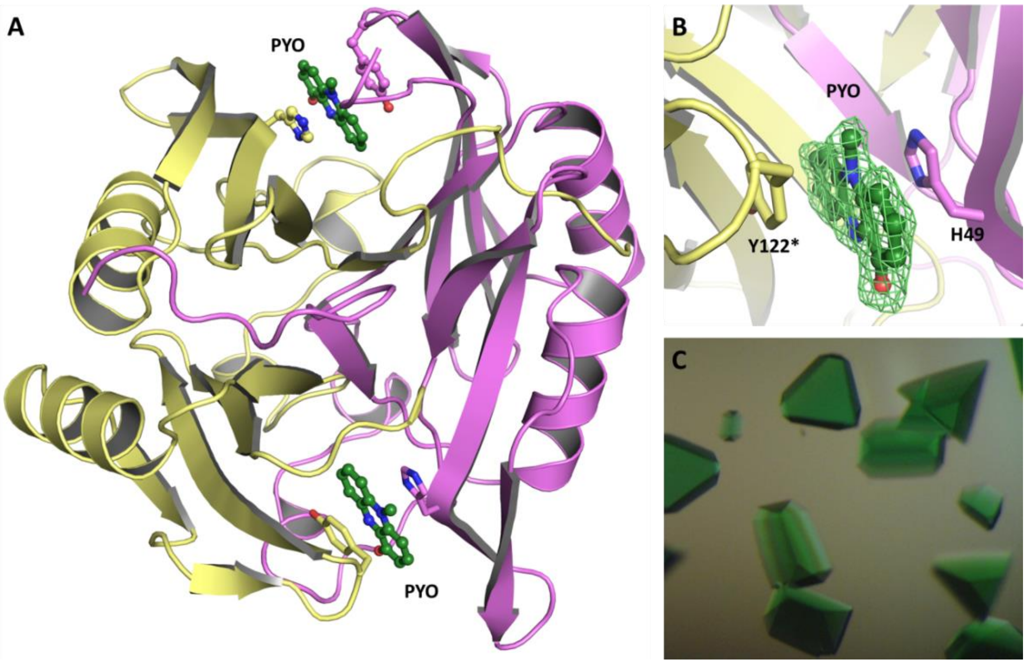


Figure 71: Crystal structures of the PA0803 homodimer with pyocyanin. A: Overall structure of the PA0803 homodimer with two pyocyanin binding sites. B: Positive m|F_{obs}|-D|F_{calc}| electron density (σ-level = 2) of pyocyanin sandwiched between H49 and Y122* by π-stacking interactions. C: PA0803 cocrystallized with pyocyanin.

Additional ITC experiments revealed that PA0803, PA1353 and PA4641 preferentially bind phenazines, and do not bind compound with one, two or four (aromatic) ring moieties. One exception to this rule is the complex antibiotic bleomycin, which possesses four aromatic moieties (though distributed over the whole molecule). Bleomycin is bound by PA0803, PA1353 and PA4641 with low micromolar affinities. However, the four rings of bleomycin are possibly not all bound simultaneously.

Functional characterization of PA0803, PA1353 and PA4641 in *E. coli* or *P. aeruginosa* using spot plate assays indicates that additional factors might play a role in pyocyanin resistance, as no differences in growth between wild type bacteria and transposon or *gain-of-function* mutants were observed. Hence, PA0803, PA1353 and PA4641 alone are not able to render *E. coli* or *P. aeruginosa* resistant against pyocyanin. One reason for the unsuccessful test may be that *E. coli* is most probably too sensitive to oxidative stress caused by pyocyanin (Hassett *et al.*, 1992). Another reason may be the absence of a membrane transporter in *E. coli* to act in concert with PA0803, PA1353 and PA4641. For this reason, *E.coli* is only able to survive in the presence of low pyocyanin concentrations (<10 µM). In contrast, *P. aeruginosa* wild type and mutants survive in the presence of more than 1 mM pyocyanin. Since pyocyanin is a signaling factor in QS of *P. aeruginosa*, the compound interferes with the transcription profile of the bacterium. For instance, the MexGHI-OpmD efflux pump, possibly a pyocyanin transporter, is upregulated by 10-fold upon pyocyanin exposure. Additionally, only single-gene transposon or *gain-of-function* mutants were tested in the spot plate assay, meaning that the other two genes are not knocked down and the resulting βαβββ-module resistance proteins could compensate for the loss of one βαβββ-module resistance protein.

7.4 Future Studies on PA0803, PA1353 and PA4641

In order to study the function of the potential pyocyanin resistance proteins PA0803, PA1353 and PA4641, clean deletion mutants of *P. aeruginosa* PAO1 have to be generated, namely ΔPA0803, ΔPA1353 and ΔPA4641, double-mutants ΔPA0803-ΔPA1353, ΔPA0803-ΔPA4641 and ΔPA1353-ΔPA4641 as well as a triple mutant ΔPA0803-ΔPA1353-ΔPA4641. These mutants should be tested in the spot plate assay with pyocyanin, because they might be more sensitive to pyocyanin than the wild type strain. Spot plates assays using bleomycin should also be conducted in order to test the hypothesis that these proteins might act as bleomycin resistance proteins, such as BLMT from *S. verticillus* (PDB entry 1NIQ, to be published). Their potential function as bleomycin resistance proteins was indicated by ITC measurements, in which the proteins bound bleomycin with low micromolar affinity.

Other methods should also be applied to test whether PA0803, PA1353 and PA4641 render *P. aeruginosa* resistant against pyocyanin. One approach is pyocyanin quantification from wild type and deletion mutant cultures under different environmental conditions, for instance in King B media which is used to enhance phenazine production. Similarly, the amount of transcribed PA0803, PA1353 and PA4641 can be tested under pyocyanin producing conditions, for instance by RNA isolation and quantification using real-time quantitative PCR. Also, transcriptome data of the deletion mutants can be recorded, or vice versa, the transcriptome of *P. aeruginosa* mutants that either do not produce pyocyanin or produce pyocyanin in large amounts. Then, the transcriptome data have to be checked for elevated or decreased levels of PA0803, PA1353 and PA4641.

Another possibility to generally identify proteins that interact with pyocyanin, protein-containing *P. aeruginosa* cell lysates can be prepared and loaded onto a column on which pyocyanin is immobilized. This column could be prepared by using biotinylated pyocyanin as a probe. Similar approaches have already been used in previous studies (Baker *et al.*, 2013). Taken together, functional characterization of the potential pyocyanin resistance proteins PA0803, PA1353 and PA4641 *in vivo* is of great importance to finally confirm their function.

7.5 Bleomycin Resistance in *P. aeruginosa*

The class II βαβββ-module resistance protein PA3127 from *P. aeruginosa* carries an additional N-acetyltransferase domain. A protein with the same domain combination is the bleomycin N-acetyltransferase BAT from *S. verticillus*. Hence, the hypothesis that PA3127 is also a bleomycin N-acetyltransferase was tested in this study. ITC experiments revealed high AcCoA and bleomycin affinities to PA3127 with K_Ds of 529 ± 33 nM and 413 ± 4 nM respectively. In an HPLC-MS based assay the conversion of AcCoA to CoA was successfully observed, however, bleomycin and acetyl-bleomycin were not detectable in this setup.

The crystal structure of PA3127 in complex with AcCoA, probably originating from *E. coli*, was determined using the *Divide & Conquer* approach. In this approach, both the βαβββ-module domain and the N-acetyltransferase domain were cloned, expressed, purified and crystallized separately in a study parallel to this work and under supervision of the author (Popp, 2013). The crystal structure of the βαβββ-module domain was successfully determined (Popp, 2013), while the N-acetyltransferase domain had to be rescreened for new crystallization conditions in this study. Finally, its structure was determined in complex with AcCoA. Then, both domains were used to determine the structure of full length PA3127 in two different crystal forms. In contrast to BAT, PA3127 is a tetramer in solution, as revealed by SAXS analysis of the protein.

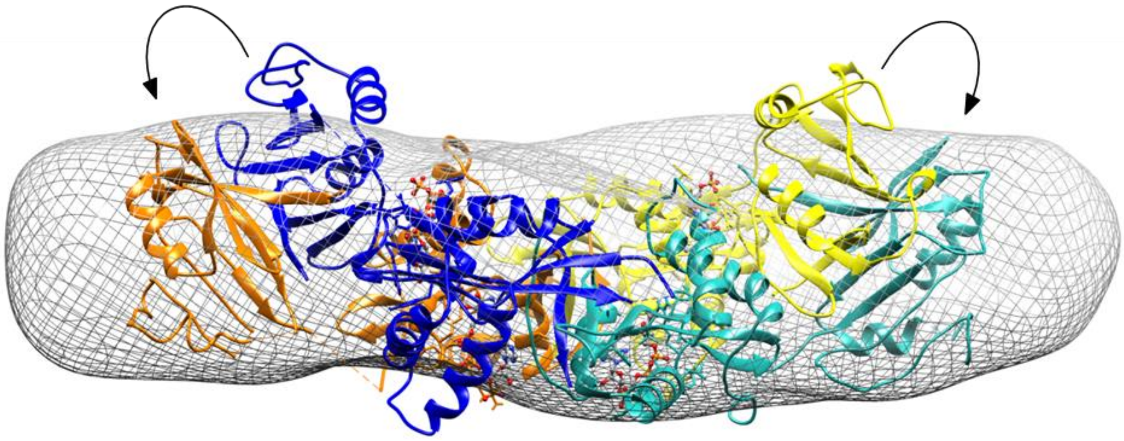


Figure 72: SAXS envelope of the PA3127 tetramer (in complex with AcCoA) which is a dimer of dimers. Tetramer formation is accomplished by dimerization of two βαβββ-module domain, whose N-acetyltransferase domains dimerize with two N-acetyltransferase domains from another PA3127 dimer. The βαβββ-module domains of the PA3127 tetramer are slightly bent out of the SAXS envelope. This conformation is caused by crystal packing. In solution, the tetramer probably relaxes an forms a plane quaternary assembly.

The SAXS envelope is of plane, rhombic shape, while the tetramer of PA3127 is slightly bent in the crystal. Thus, it is rather difficult to speculate about the bleomycin binding mode to PA3127. Most probably, the metal-free bleomycin stretches along both domains, with its bis-thiazole moiety bound between Y31, Y34 and W96 of the βαβββ-module domain, and the primary amine bound in the proximity of the acetyl group of AcCoA. Efficient transfer of the acetyl group of AcCoA to bleomycin probably occurs by direct nucleophilic attack of the primary amine of bleomycin to AcCoA, as suggested for other members of the N-acetyltransferase superfamily. The only candidate residue to abstract a proton from the primary amine of bleomycin in order to turn it into an efficient nucleophile is E205. However, this hypothesis can probably only be proven by a crystal structure of a PA3127-bleomycin-AcCoA complex, or possibly by an inactive mutant of PA3127 that could either be tested in an activity or a spot plate assay. A potential mutation candidate is residue E205 which might be able to abstact a proton from the primary amino group of bleomycin, thereby rendering it an active nucleophile.

7.6 Future Studies on PA3127 and Its Operon

Future studies on PA3127 should include the cocrystallization of the protein with bleomycin or AcCoA and bleomycin to determine the bleomycin binding mode and the orientation of the substrates towards each other. Furthermore, kinetic assays should be performed to characterize the protein’s catalytic activity. For this purpose, either the HPLC-MS assay setup has to be optimized, or the assay used for BAT from *S. verticillus* has to be adapted to PA3127. Still, the initial hypothesis of PA3127 being a bleomycin N-acetyltransferase could be confirmed by combining ITC, X-ray crystallography, SAXS and *in vivo* microbiological assays.

PART II

Structural Insights into Mycobacterial Ergothioneine Biosynthesis

8 Summary

Ergothioneine is a sulfur derivative of histidine betaine mainly produced by Actinobacteria and Basidiomycota. It has versatile functions as an antioxidant, metal chelator and it is involved in the biosynthetic reaction of lincomycin A. Humans are not able to biosynthesize ergothioneine, but resorb it from their diet and store it using a specific ergothioneine transporter ETT. Ergothioneine and its transporter are important players in human physiology. Disorders in ergothioneine homeostasis are for instance involved in Crohn's disease and induce oxidative stress.

In order to investigate mycobacterial ergothioneine biosynthesis, the SAM-dependent methyltransferase EgtD, the sulfoxide synthase EgtB and the Ntn hydrolase EgtC were examined using X-ray crystallography, biophysical and biochemical approaches in collaboration with Prof. Dr. Florian P. Seebeck from the University of Basel, Switzerland. The apo structures of EgtC and EgtD have been determined in a study prior to this work.

In this study, substrate complexes of EgtD with N^{α},N^{α}-dimethyl-histidine and SAH were obtained by X-ray crystallography, and kinetic data of EgtD were recorded. Using these data, an EgtD variant specific for tryptophan was engineered, assayed and crystallized with tryptophan and SAH. The data do not only characterize the first and rate-determining step in ergothioneine biosynthesis, but for the first time they describe aromatic amino acid betaine biosynthesis and identify other aromatic amino acid betaine-producing enzymes. Furthermore, the Ntn hydrolase EgtC was cocrystallized with glutamine and its substrate γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide, resolving the stereochemistry at the sulfoxide position of the substrate. On the basis of these structures, a potential model of the active conformation and binding mode of enzyme and substrate was proposed. Additionally, the crystal structure of the sulfoxide synthase EgtB was determined via sulfur SAD. The enzyme represents a novel class of sulfoxide synthases. A quaternary complex of EgtB with N^{α},N^{α}-dimethyl-histidine and γ-glutamyl-cysteine reveals the substrates as direct ligands to the iron metal center and allows deriving a mechanism that is in agreement with the sulfoxide stereochemistry determined in the EgtC substrate structure. Finally, inhibitor design for EgtD identified α-chloro-histidine as the most potent inhibitor with a K_I of 2.5 µM in a coupled EgtD assay, and lays the basis for further inhibitor design.

9 Introduction

Bacteria, archaea and fungi produce and metabolize a huge variety of substances, reaching from chemically rather simple compounds such as fumarate to complex antibiotics such as bleomycin. Even if some of these compounds have been known for more than hundred years, their biosynthesis and molecular mode of action may not have been elucidated yet. This might be due to the fact that their biosynthetic proteins and the respective genes have not been identified. In recent years, the availability and speed of genome sequencing have increased and bioinformatic tools can easily be applied to sequenced genomes. Using this approach, initial information on bacterial and archaeal genomes can be gained and the functional prediction of the respective genes can be accomplished. Additionally, X-ray crystallography is a powerful tool for the identification a molecular in-depth characterization of the enzymes catalyzing the biosynthetic reactions. On the one hand, this is important to understand pathogens and their action. On the other hand, these metabolites may be potential therapeutics, or their biosynthesis may be potential drug targets.

One example for such a compound is the small metabolite N^{α},N^{α},N^{α}-trimethyl-2-thio-histidine (Figure 73) which was isolated from ergot fungus *Claviceps purpurea* in 1909 and therefore named ergothioneine (Tanret, 1909). Ergothioneine is a betaine derivative of 2-thio-histidine. In principal, ergothioneine exists in two tautomeric forms, a thiol and a thione/thioketone form. Under physiological conditions, the thione form is preferred over the thiol form (Akanmu *et al.*, 1991), which is a characteristic property of ergothioneine.

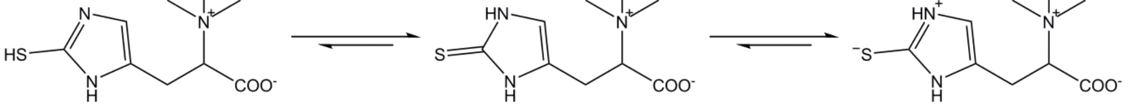


Figure 73: Tautomeric thiol- (left), thione/thioketone (middle) and the mesomeric thioenol forms of ergothioneine. The thioketone form (middle) is preferred under physiological conditions.

Additionally, the pK_a of the thioimidazole group of ergothioneine is unusually high (>10) compared to most alkylthiols (Hartman, 1990).

9.1 Properties and Functions of Ergothioneine in Health and Disease

After being discovered in fungi, ergothioneine was detected in microorganisms, plants and mammals (Salt, 1931; Genghof *et al.*, 1956; Melville & Eich, 1956). However, humans are not able to biosynthesize ergothioneine, but absorb it with their diet. Although the precise function of ergothioneine has to be confirmed *in vivo* and no proof has been found yet that ergothioneine is an essential metabolite in humans, humans possess an ergothioneine-specific transporter ETT (Gründemann *et al.*, 2005). The ETT inhomogenously distributes ergothioneine to accumulate it in concentrations of 100 µM to 2 mM in tissues that are exposed to oxidative stress. In fact, ergothioneine is mainly found in erythrocytes, eye lense, kidney, liver and seminal fluid (Hunter, 1928; Salt, 1931; Leone & Mann, 1951; Melville *et al.*, 1954). Silencing of the ETT-encoding gene inhibits ergothioneine uptake, indicating that there is no alternative way of transport (Paul & Snyder, 2010). Further observations in bacterial, fungal and animal cells indicate that ergothioneine is an important player in oxidative stress response and physiology.

In contrast to glutathione, which is readily oxidized by Cu(II) and Fe(II), ergothioneine is stable against oxidation by divalent transition metal cations. Two ergothioneine molecules usually act as chelators of one Cu(II), Fe(II), Ni(II), Cd(II), Zn(II), Hg(II) or Co(II) ion (Hanlon, 1971; Motohashi *et al.*, 1974, 1976). More generally, ergothioneine is neither prone to oxidation nor to disulfide bond formation. Thus, it is rather present in its reduced form. This finding clearly delimits it from GSH or mycothiol (MSH). The formation of oxidized ergothioneine only occurs under strong acidic conditions (Hand & Honek, 2005). A reason for ergothioneine preferring the reduced state is the unique redox potential of its thiol–disulfide couple (−0.06 V) in contrast to the ones of other naturally occurring thiols such as glutathione (−0.2 and −0.32 V) (Jocelyn, 1958).

Furthermore, mitochondrial DNA damage was found to increase in response to ETT silencing when cells where exposed to H_2O_2 (Paul & Snyder, 2010). Hence, ergothioneine might be able to protect mitochondrial DNA from damage induced by reactive oxygen species (ROS) generated by the electron transport chain. *In vitro* experiments indeed demonstrate that ergothioneine is able to scavenge ROS such as peroxonitrite ($ONOO^-$), hypochlorous acid (HOCl) or hydroxyl radicals ($^\bullet OH$) *in vitro* and to protect cells from UV- and γ-radiation damage (Motohashi & Mori, 1986; Akanmu *et al.*, 1991; Franzoni *et al.*, 2006). Since ergothioneine is able to absorb light in the UV range, it was suggested to block UV-induced ROS generation and damage. The cell viability of keratinocytes significantly decreased upon exposure to UV light, but was increased upon addition of ergothioneine (Botta *et al.*, 2008; Markova *et al.*, 2009). However, this hypothesis is controversially discussed in the field of ergothioneine research, since recent findings indicate that ergothioneine is required for the production of spores in *N. crassa*, but does not protect against 254 nm UV-induced mutagenesis or kill (Bello *et al.*, 2014). An overview on the proposed functions of ergothioneine and ETT is given in Figure 74.

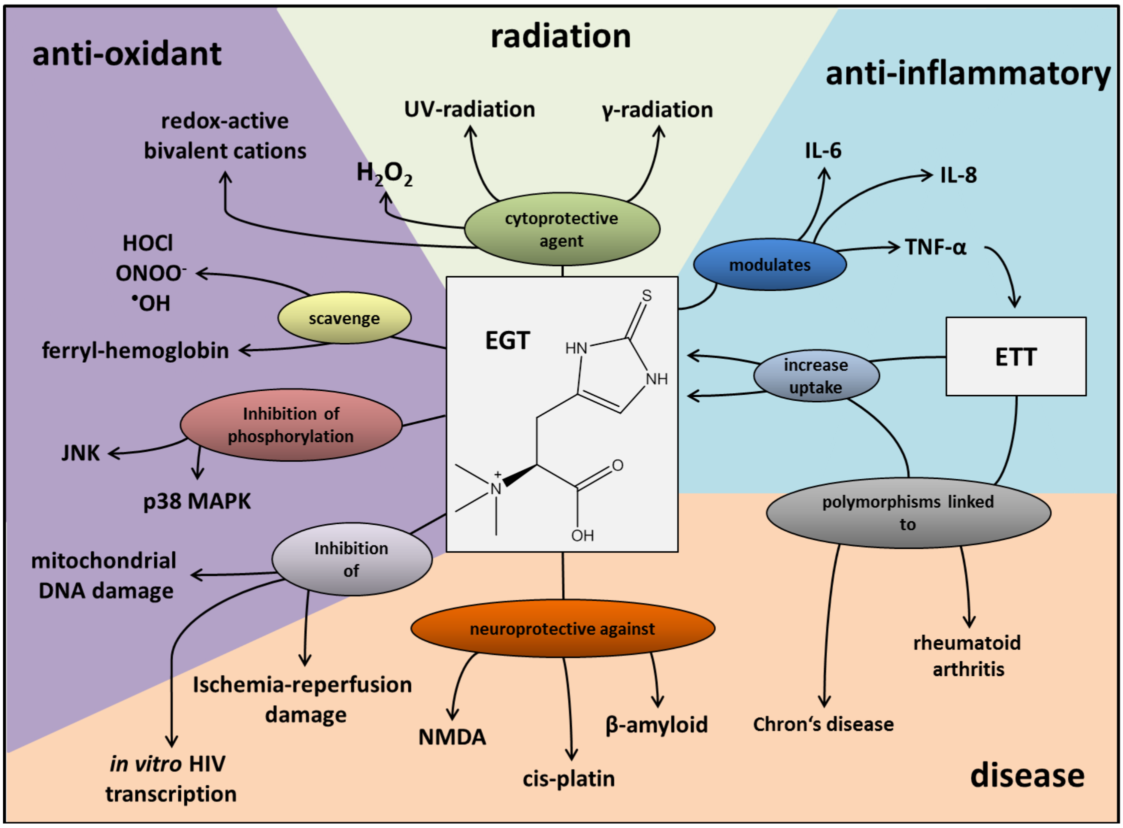


Figure 74: Proposed functions of Ergothioneine (EGT). Ergothioneine and its transporter protein ETT might be involved in several cellular processes in health and disease in ergothioneine producers as well as in non-producers, for instance humans. Figure adapted from Cheah & Halliwell, 2012.

The roles of ergothioneine and its transporter in inflammation are not fully understood yet. Mutations of the ETT gene are involved in chronic inflammatory diseases including rheumatoid arthritis (Taubert *et al.*, 2009) and Crohn's disease (Peltekova *et al.*, 2004). Patients suffering from rheumatoid arthritis and Crohn's disease have elevated ergothioneine levels in erythrocytes and monocytes due to gain-of-function mutations in the ETT (Peltekova *et al.*, 2004). Silencing of ETT expression in erythrocytes in vitro impairs ergothioneine uptake, induces oxidative stress in HeLa cells and has been shown to be pro-apoptotic (Nakamura *et al.*, 2007; Paul & Snyder, 2010). Hence, increased ergothioneine concentrations under these chronic inflammatory conditions are postulated to promote immune cell survival (Taubert *et al.*, 2009; Paul & Snyder, 2010). Additionally, ETT knock-out mice were shown to be susceptible to intestinal inflammation, suggesting that ergothioneine may counteract intestinal inflammation. In support of this hypothesis, ergothioneine has been shown to inhibit the TNFα-mediated increase of IL-8 in epithelial cell cultures (Rahman *et al.*, 2003). Silencing of the ETT in *Caenorhabditis elegans* decreases the life span of this nematode and increases its susceptibility to oxidative stress (Cheah *et al.*, 2013). Additionally, ergothioneine protects against palmitic acid-induced apoptosis through activation of p38 and JNK kinase signaling pathways and inhibition of IL-6 expression (Laurenza *et al.*, 2008).

Taken together, these observations and the presence of a specific transporter for ergothioneine suggest that ergothioneine is a beneficial metabolite. This strongly supports the theory that ergothioneine homeostasis is important for human physiology and health. Hence, it is important to study its role and its biosynthesis. The latter, however cannot be accomplished in humans, but only in organisms that biosynthesize ergothioneine, such as the model organism *Mycobacterium smegmatis*.

9.2 Discovery and *in vitro* Reconstitution of Ergothioneine Biosynthesis

Although ergothioneine was discovered more than hundred years ago, its exact biosynthetic pathway and the genes encoding ergothioneine-biosynthetic enzymes remained unidentified for decades, because research focused more on the occurrence, isolation and on functional characterization of ergothioneine.

Early studies on mycobacterial isolates confirmed ergothioneine biosynthesis for many species belonging to this genus, such as *Mycobacterium tuberculosis* or *M. smegmatis* (Genghof & Van Damme, 1964, 1968). Today, ergothioneine biosynthesis is known to occur in Actinobacteria, Cyanobacteria, Pezizomycotina and Basidiomycota and in numerous α-,β-,γ- and δ-Proteobacteria and Bacteroidetes (Genghof, 1970; Seebeck, 2010; Pfeiffer *et al.*, 2011). Historically, first attempts to identify the precursor of ergothioneine biosynthesis were made in 1950. Feeding experiments in *Neurospora crassa* and

Claviceps purpurea demonstrated that radioactively labeled histidine was fully and intactly incorporated into ergothioneine (Melville *et al.*, 1957; Heath & Wildy, 1958). Hence, histidine was proven to be the precursor amino acid of ergothioneine and in the same study, cysteine was concluded to provide the sulfur atom for ergothioneine (Melville *et al.*, 1957). On the basis of these results, either 2-thiol-histidine or N^{α},N^{α},N^{α}-trimethyl-histidine was proposed to be intermediates of ergothioneine biosynthesis. The latter was detected in *N. crassa* mycelium and postulated to be the first intermediate in ergothioneine biosynthesis (Askari & Melville, 1962). At the same time, evidence for a single enzyme catalyzing all three methylation reactions of histidine to N^{α},N^{α},N^{α}-trimethyl-histidine was found (Ishikawa & Melville, 1970), and S-adenosyl-methionine (SAM) was identified as the methyl group donor (McManus, 1962; Ishikawa & Melville, 1970). Later, another intermediate of ergothioneine biosynthesis, N^{α},N^{α},N^{α}-trimethyl-histidinyl-cysteine sulfoxide, was isolated from cell-free *N. crassa* extracts (Ishikawa *et al.*, 1974).

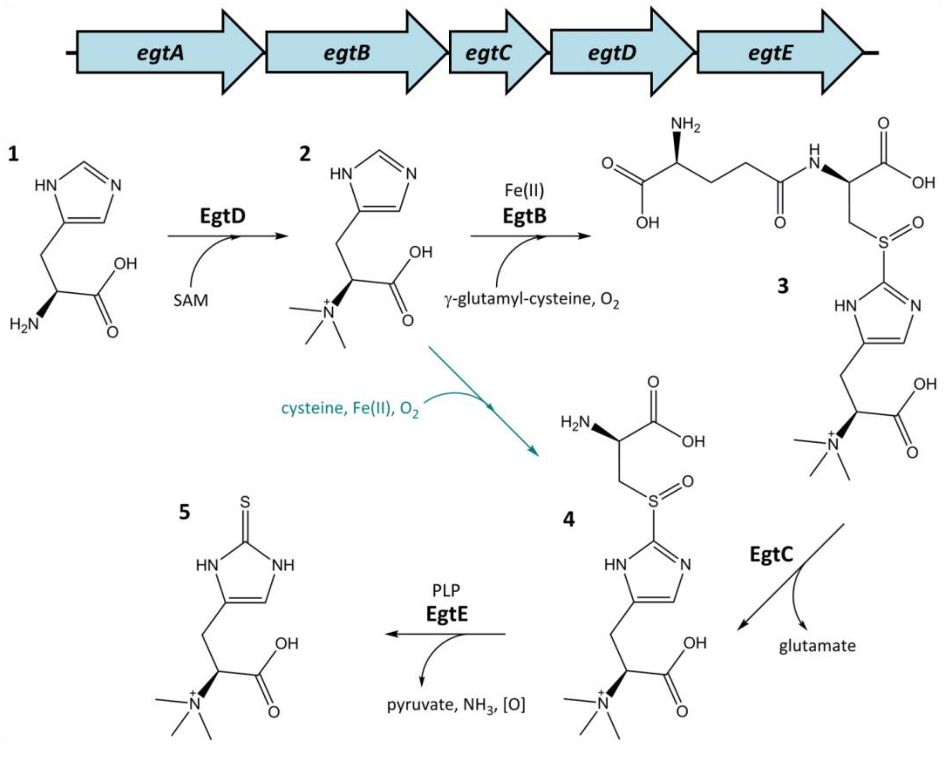


Figure 75: Ergothioneine biosynthesis pathway as proposed by Seebeck, 2010. The ergothioneine-biosynthetic gene cluster from *M. smegmatis* contains five genes *egtABCDE*. Ergothioneine is produced from histidine via a trimethylation reaction catalyzed by the SAM-dependent methyltransferase EgtD. The non-heme Fe(II) enzyme EgtB catalyzes the formation of an isopeptide intermediate which is hydrolyzed by the Ntn hydrolase EgtC and converted to ergothioneine by the pyridoxal 5-phosphate (PLP)-dependent lyase EgtE. γ-glutamyl-cysteine is provided by EgtA. A previously suggested pathway in *N. crassa* (teal) (Ishikawa & Melville, 1970; Ishikawa *et al.*, 1974) turned out to be correct (Hu *et al.*, 2014). Figure adapted from Seebeck, 2010.

The production of this intermediate occurs only in the presence of Fe(II), probably as a metal cofactor of an enzyme, and O_2 (Ishikawa *et al.*, 1974). On the basis of these results, a more detailed path of ergothioneine biosynthesis was proposed (Ishikawa *et al.*, 1974). However, it took more than 35 years until ergothioneine biosynthesis was reconstituted *in vitro* and all reaction steps were identified in mycobacterial ergothioneine biosynthesis (Seebeck, 2010). Since it was known that a SAM-dependent methyltransferase, an Fe(II)-dependent oxidase and a pyridoxal 5-phosphate (PLP)-dependent lyase are probably involved in ergothioneine biosynthesis (Ishikawa & Melville, 1970; Ishikawa *et al.*, 1974), the mycobacterial ergothioneine biosynthesis genes could be identified by a bioinformatic approach. For this purpose, the genomes of ergothioneine producers *N. crassa* and *Mycobacterium avium* were searched for methyltransferases encoded in both genomes, but not encoded in the genomes of *Bacillus subtilis* and *E. coli*, which do not produce ergothioneine. This approach identified ten methyltransferase in the genome of *M. avium,* but only one of them encoded in an operon with a PLP-dependent lyase. Enzymes from this operon were cloned and the intermediates and final product of ergothioneine biosynthesis were analyzed via HPLC and Electrospray Ionization Mass Spectrometry (ESI-MS) (Seebeck, 2010). As depicted in Figure 75, mycobacterial ergothioneine (**5**) biosynthesis starts from histidine (**1**) which is N^{α}-trimethylated by the SAM-dependent methyltransferase EgtD. The reaction product $N^{\alpha},N^{\alpha},N^{\alpha}$-trimethyl-histidine (**2**), is passed on to the non-heme Fe(II) enzyme EgtB. This enzyme catalyzes the O_2-dependent carbon-sulfur bond formation and oxidation of $N^{\alpha},N^{\alpha},N^{\alpha}$-trimethyl-histidine and γ-glutamyl-cysteine (produced by EgtA). The product of the EgtB-catalyzed reaction, γ-glutamyl-$N^{\alpha},N^{\alpha},N^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide (QHC, **3**) is then hydrolyzed to glutamate and $N^{\alpha},N^{\alpha},N^{\alpha}$-trimethyl-histidinyl-cysteine sulfoxide (**4**) by the Ntn amidohydrolase EgtC. In the final step, the PLP-dependent lyase EgtE catalyzes the formation of ergothioneine.

9.3 Characteristic Features of Ergothioneine Biosynthesis

9.3.1 Genetics of Ergothioneine Biosynthesis

In Mycobacteria, the enzymes EgtABCDE catalyze the biosynthesis of ergothioneine. These enzymes are encoded in the mycobacterial ergothioneine biosynthesis gene cluster *egtABCDE* (Seebeck, 2010). The analysis of more than 2500 ergothioneine-biosynthetic gene clusters revealed that the five-gene ergothioneine cluster is a specific trait of Actinobacteria (Jones *et al.*, 2014). Actinobacteria might have passed on their ergothioneine biosynthesis genes to δ-Proteobacteria (e. g. *Myxococcus xanthus*). However, some ergothioneine producers lack EgtA, EgtC and/or EgtE, the latter being exclusively present in Actinobacteria. For instance Cyanobacteria lack EgtACE, but still produce high levels of ergothioneine (Pfeiffer *et al.*, 2011). Possibly, EgtA, EgtC and EgtE can be complemented by enzymes encoded elsewhere in the genome (Jones *et al.*, 2014).

Recently, also the first fungal ergothioneine biosynthesis gene *Ncegt-1* from *N. crassa* (Bello *et al.*, 2012) has been identified. The resulting protein, NcEgt-1 is probably the result of a gene fusion event between ancestral *egtB* and *egtD* (Bello *et al.*, 2012). Fungi do not produce EgtA, EgtC and EgtE, either. This finding indicates that EgtB and EgtD are the key enzymes of ergothioneine biosynthesis (as predicted earlier by (Seebeck, 2010). Taken together these findings strongly support that the (myco)bacterial ergothioneine biosynthesis route is probably distinct from the fungal pathway (Figure 75, Figure 77A).

9.3.2 Glycine Betaine and Aromatic Amino Acid Betaine Biosynthesis

Inside the cell, amino acids can be modified to fulfill other functions than being the building blocks of proteins. As seen in ergothioneine biosynthesis, they can be modified to be amino acid betaines, in this case N^{α},N^{α},N^{α}-trimethyl-histidine (Figure 76A). So far, the most studied amino acid betaine is glycine betaine. Glycine betaine is biosynthesized from its precursors choline or even phosphoethanolamine in a multi-enzyme pathway (Figure 76B). In its primary role, it is thought to act as a potent osmoprotectant (Wargo, 2013). However, the biosynthesis pathway of glycine betaine is hardly adaptable to aromatic amino acid betaines, because they are directly synthesized from the corresponding amino acid, and not from choline (Figure 76). These betaines constitute a ubiquitous class of natural compounds. Aromatic amino acids betaines and their derivatives of all four aromatic amino acids phenylalanine, histidine, tryptophan and tyrosine have been discovered. They are present in plants and fungi, for instance as phenylalanine betaine (N^{α},N^{α},N^{α}-trimethyl-phenylalanine), sticticine (dihydroxy-N^{α},N^{α},N^{α}-trimethyl-tyrosine) or hypaphorine (N^{α},N^{α},N^{α}-trimethyl-tryptophan) (Carollo *et al.*, 2010; Mollica *et al.*, 2012).

Hence, aromatic amino acid betaines play an important role in life and are part of our daily nutrition.

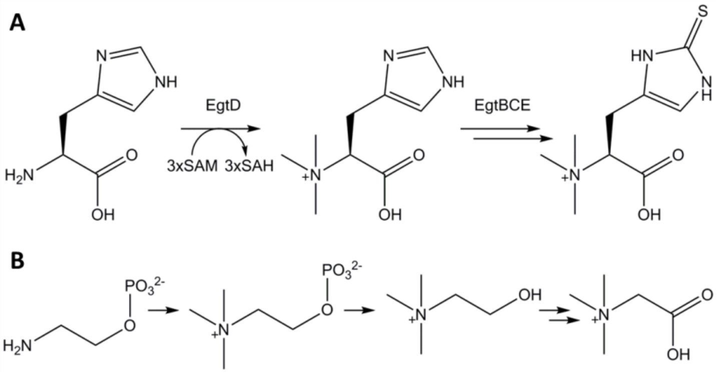


Figure 76: Comparison of glycine and histidine betaine biosynthesis. A: In ergothioneine biosynthesis, only EgtD is sufficient to produce N^{α},N^{α},N^{α}-trimethyl-histidine from histidine. Then, EgtB, EgtC and EgtE catalyze all further steps of ergothioneine biosynthesis. B: Glycine betaine is produced from phosphoethanolamine in *P. falciparum* (Lee *et al.*, 2012), but at least from choline, not from its corresponding amino acid glycine, in a multi-enzyme pathway. Hence, aromatic amino acid betaine biosynthesis is distinct from glycine betaine biosynthesis.

N^{α},N^{α},N^{α}-trimethyl-histidine is the aromatic amino acid betaine derivatives which is produced by EgtD in the first step of ergothioneine biosynthesis. Only this one enzyme is sufficient to produce N^{α},N^{α},N^{α}-trimethyl-histidine from histidine, while several biosynthetic steps are necessary for the production of glycine betaine. Hence, EgtD structure elucidation promotes the understanding of the first step of ergothioneine biosynthesis, as well as the structural basis of aromatic amino acid betaine biosynthesis. The structure of EgtD will help to identify related enzymes that are able to produce other aromatic amino acid betaines.

9.3.3 Sulfur Donors and Sulfoxide Stereochemistry

As mentioned earlier, mycobacterial ergothioneine biosynthesis is catalyzed by five proteins EgtABCDE (Seebeck, 2010), while fungi probably use NcEgt1 homologs to produce ergothioneine (Bello *et al.*, 2012; Hu *et al.*, 2014). The corresponding gene *Ncegt-1* is most likely the result of a gene fusion event between ancestral *egtB* and *egtD* (Bello *et al.*, 2012). Hence, Nc-Egt-1 catalyzes the first two steps in *N. crassa* ergothioneine biosynthesis and therefore converts histidine (**1**) to N^{α},N^{α},N^{α}-trimethyl-histidinyl-cysteine sulfoxide (**4**, Figure 77B, Bello *et al.*, 2012). Both NcEgt-1 and EgtB need sulfur donor molecules to catalyze their reactions. NcEgt-1 has a 62-fold higher specificity for cysteine as a sulfur donor relative to γ-glutamyl-cysteine (Hu *et al.*, 2014), while mycobacterial EgtB use γ-glutamyl-cysteine (Ishikawa *et al.*, 1974; Seebeck, 2010; Bello *et al.*, 2012). Since *Ncegt-1* homologs are distributed across all fungal genomes except for *Saccharomycotina* species (Jones *et al.*, 2014), this might indicate that all fungi rather use cysteine than γ-glutamyl-cysteine as a sulfur donor (Bello *et al.*, 2012). This switch in substrate specificity eliminates the competition between glutathione (needs γ-glutamyl-cysteine as a precursor) and ergothioneine biosynthesis in fungal species (needs cysteine as a precursor) (Hu *et al.*, 2014). In *N. crassa*, the last step of ergothioneine biosynthesis is catalyzed by the C-S lyase NCU11365 recently identified by genome mining; subsequently *N. crassa* ergothioneine biosynthesis was also characterized *in vitro* (Hu *et al.*, 2014). The two routes of biosynthesis in Mycobacteria and in fungi are depicted in Figure 77.

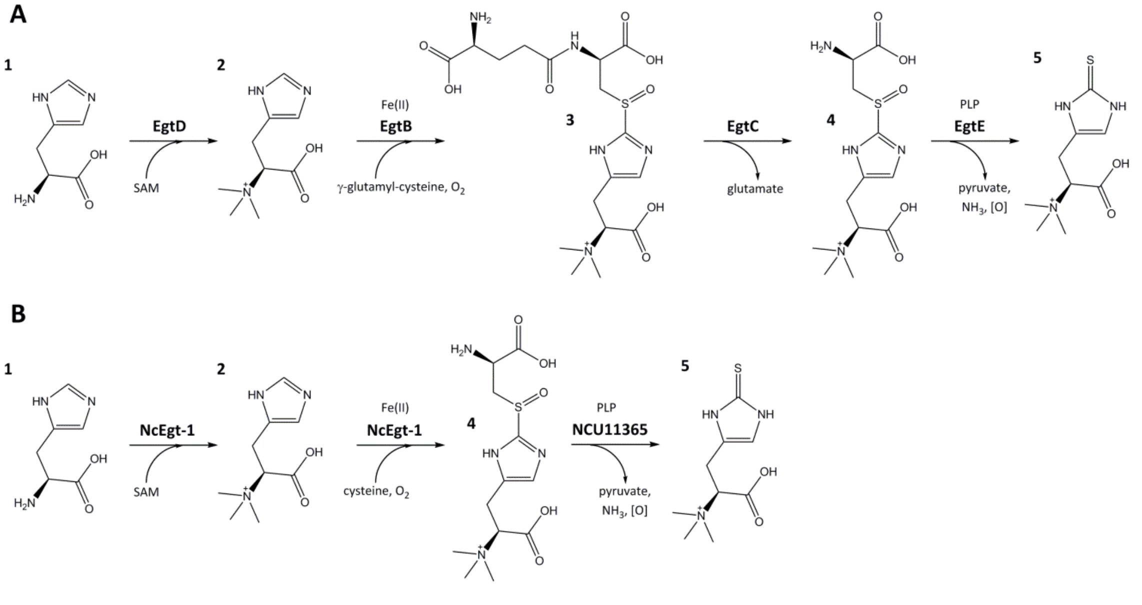


Figure 77: Comparison of mycobacterial and fungal ergothioneine biosynthesis. A: Mycobacterial ergothioneine biosynthesis 5-step pathway as proposed by Seebeck, 2010. Mycobacteria use γ-glutamyl-cysteine produced by EgtA as a sulfur donor molecule. B: Ergothioneine biosynthesis in *N. crassa* (Bello *et al.*, 2012; Hu *et al.*, 2014). The first two steps are catalyzed by NcEgt-1, a fusion protein of EgtB and EgtC. The C-S lyase NCU11365 from *N. crassa* then catalyzes the final formation of ergothioneine (Hu *et al.*, 2014).

Although fungi and Mycobacteria use different sulfur donor molecules, cysteine and γ-glutamyl-cysteine, respectively, they both employ EgtB (or the EgtB domain of NcEgt-1 homologs) to catalyze the central carbon-sulfur bond formation and oxidation step in ergothioneine biosynthesis. Interestingly, this step involves the selective formation of one stereoisomer of QHC (**3**) or N^{α},N^{α},N^{α}-trimethyl-histidinyl-cysteine sulfoxide (**4**) at the respective sulfur atom. However, it is not yet clear how the correct stereoisomers of QHC or N^{α},N^{α},N^{α}-trimethyl-histidinyl-cysteine sulfoxide are formed, and which of the possible stereoisomers is formed in the reaction. Characterization of the product stereochemistry is especially difficult because EgtB represents a novel type of C-S bond-forming enzymes and no structure of an enzyme from this class is available (Seebeck, 2010; Hu *et al.*, 2014). The protein is composed of an N-terminal α-helical DinB domain and a C-Lectin (CLec)-like domain that is characterized by a very low amount of secondary structure elements. Hence, the exact three-dimensional structure as well as the domain arrangement and the atomic detail of the active center remain unknown. Thus, the binding mechanism and positioning of the substrates in the active center are unpredictable and have to be elucidated. In order to draw conclusions on the enzymatic mechanism employed by EgtB and on the product stereochemistry, it is essential to obtain structural information of mycobacterial EgtB and EgtB complexes with its substrates N^{α},N^{α},N^{α}-trimethyl-histidine and γ-glutamyl-cysteine. Additionally, the structure of this enzyme would provide detailed insight into a new type of reaction mechanism, as the structure of EgtB would be the first structure of a DinB-CLec protein and the catalytic residues are unknown.

The product of the EgtB-catalyzed reaction, QHC (**3**), is the substrate for the hydrolysis catalyzed by the Ntn amidohydrolase EgtC. The apo structure of EgtC has been determined earlier via Se SAD (Vit, 2010), but no substrate complexes have yet been obtained. In order to prove the stereochemistry of QHC, the co-crystallization of an inactive EgtC variant in complex with QHC is crucial, since previous attempts to crystallize QHC alone failed. Furthermore, there are no EgtC homologs and very few Ntn hydrolase structures deposited in the PDB. The structural characterization of EgtC will generate a deeper understanding of the class of Ntn amidohydrolases.

9.4 Ergothioneine Biosynthesis and EgtD as Therapeutic Targets?

Ergothioneine is biosynthesized by Actinobacteria, which includes pathogens such as *M. tuberculosis*. This fact classifies ergothioneine biosynthesis as a therapeutic target, for instance in tuberculosis treatments. One reason is that the compound seems to fulfill distinct functions in Mycobacteria, different from the functions of other antioxidant such as mycothiol. This antioxidant consists of an N^{α}-acetylated cysteine, a glucosamine moiety and an inositol group. In contrast to ergothioneine, mycothiol is not secreted by *M. smegmatis* (Sao Emani *et al.*, 2013). This explains why mycothiol cannot fully compensate for loss of ergothioneine in Mycobacteria (Ta *et al.*, 2011). Interestingly, ergothioneine was recently reported to act as a template molecule in the biosynthesis of lincomycin A (Zhao *et al.*, 2015), a sulfur-containing lincosamide antibiotic that has been widely used in veterinary medicine. Thus, ergothioneine is involved in an enzymatic reaction and biosynthesis pathway itself, indicating that more of these reactions might be discovered in the future. For these reason, it might be additionally beneficial to inhibit ergothioneine biosynthesis. However, the genes that encode the ergothioneine biosynthesis proteins were predicted not to be essential for growth of the *M. tuberculosis* H37Rv laboratory strain (Griffin *et al.*, 2011).

Generally speaking, the first step in a pathway is often a suitable target for drug development. In ergothioneine biosynthesis, this first step - the conversion of histidine to N^{α},N^{α},N^{α}-trimethyl-histidine by EgtD - is essential to initiate further ergothioneine biosynthesis steps, as the following enzymes EgtBCE only used trimethylated substrates. EgtD has no homologs in humans and was demonstrated to be an essential enzyme in ergothioneine biosynthesis, as it is conserved in all ergothioneine producers (Jones *et al.*, 2014). Elimination of EgtD from *M. smegmatis* abolishes ergothioneine biosynthesis (Sao Emani *et al.*, 2013). These facts further confirm its essential role in ergothioneine biosynthesis. Furthermore, the purification protocol and the crystal structure of EgtD are already available (Vit, 2010; Vit *et al.*, 2014, 2015). Hence, the active center of EgtD can easily be assessed to design potent EgtD inhibitors which can then be tested in a biochemical assay and eventually be cocrystallized or soaked into EgtD.

10 Research Objectives

The biosynthesis operon of mycobacterial ergothioneine, a betaine derivative of 2-thiohistidine, has only been reconstituted *in vitro* in 2010 (Seebeck, 2010). In a study prior to this work, the apo structures of the SAM-dependent N^{α},N^{α},N^{α}-trimethyltransferase EgtD and the Ntn hydrolase EgtC were determined using selenium SAD in X-ray crystallography (Vit, 2010). Additionally, substrate binding to EgtD was characterized by isothermal titration calorimetry (ITC).

First, this study aims at structure determination of substrate complexes of EgtD from *M. smegmatis*. These complexes deepen the understanding of the first and rate-determining step in ergothioneine biosynthesis catalyzed by EgtD. Additionally, they lay the structural basis of aromatic amino acid betaine biosynthesis, since EgtD is the first enzyme from this family that is structurally characterized. Additionally, EgtD is an interesting therapeutic target, since removal of *egtD* from *M. smegmatis* abolishes ergothioneine biosynthesis. Crystal structures of EgtD inhibitor complexes were therefore determined. The Ntn hydrolase EgtC catalyzes the third step in ergothioneine biosynthesis. An inactive variant of EgtC was cocrystallized with the isopeptide substrate γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide (QHC) to unveil the stereochemistry at the sulfoxide position.

Furthermore, the structure of the non-heme Fe(II) enzyme EgtB from *M. thermoresistibile* was determined. This enzyme catalyzes the oxidative coupling of N^{α},N^{α},N^{α}-trimethyl-histidine and γ-glutamyl-cysteine to N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide. Since there is no structure of a two-domain CLec-like protein available in the PDB, EgtB would be the first enzyme of this kind. In addition to that, structural and kinetic characterization of EgtB and substrate complexes will answer the question which residues form the Fe(II) complex or where and how the substrates bind to the enzyme. The mode of substrate binding is crucial for the biosynthesis of the correct stereoisomer of N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide.

These structures elucidate the path of ergothioneine biosynthesis, give insights into novel protein families, and they will pave the way to novel therapeutics and enzyme engineering approaches.

11 List of Publications

In this section, my own publications associated with mycobacterial ergothioneine biosynthesis are listed. In order to understand the molecular basis of ergothioneine biosynthesis, the ergothioneine biosynthesis proteins EgtB, EgtC and EgtD were investigated in collaboration with Prof. Dr. Florian P. Seebeck at the University of Basel in Switzerland. The apo structures of the N-terminal nucleophilce (Ntn) amidohydrolase EgtC and the methyltransferase EgtD were determined earlier (Vit, 2010) and parts of the results have been published (Vit *et al.*, 2014). The ITC data on EgtD as well as the apo structures of EgtD and EgtC were obtained during my master thesis (Vit, 2010). These results form the basis for the experiments performed during my PhD studies. In this study, structure elucidation of enzyme substrate complexes of EgtC and EgtD was pursued. As insights into the novel sulfoxide synthase reaction of EgtB are of high impact to the field, this novel enzyme was investigated exclusively in this study. Additionally, inhibitor complexes of EgtD were obtained.

Publication A

Allegra Vit*, Laëtitia Misson*, Wulf Blankenfeldt and Florian P. Seebeck[§]. **Ergothioneine Biosynthetic Methyltransferase EgtD Reveals the Structural Basis of Aromatic Amino Acid Betaine Biosynthesis**. *ChemBioChem*. **2015** Jan 2; 16(1): 119-125.

Publication B

Kristina V. Goncharenko*, Allegra Vit*, Wulf Blankenfeldt[§] and Florian P. Seebeck[§]. **Structure of the Sulfoxide Synthase EgtB from the Ergothioneine Biosynthetic Pathway**. *Angew. Chem. Int. Ed. Engl*. **2015** Feb 23; 54(9): 2821-2824.

Manuscripts in Preparation

Allegra Vit, Gabriel T. Mashabela, Wulf Blankenfeldt and Florian P. Seebeck[§]. **Structure of the Ergothioneine Biosynthetic Amidohydrolase EgtC**.

Laëtitia Misson, Allegra Vit, Reto Brun, Sanja Kostic, Wulf Blankenfeldt, Florian P. Seebeck[§]. **Inhibiton of the Ergothioneine Biosynthetic Methyltransferase EgtD.**

12 Results

12.1 Ergothioneine Biosynthetic Methyltransferase EgtD Reveals the Structural Basis of Aromatic Amino Acid Betaine Biosynthesis

Publication A

Allegra Vit*, Laëtitia Misson*, Wulf Blankenfeldt and Florian P. Seebeck§. Ergothioneine Biosynthetic Methyltransferase EgtD Reveals the Structural Basis of Aromatic Amino Acid Betaine Biosynthesis. *ChemBioChem*. **2015** Jan 2; 16(1): 119-25.

*Both authors contributed equally to this work.
§Corresponding author.

Author Contribution

I cloned, produced and purified $\mathrm{EgtD_{wt}}$ and produced and purified HyoA for crystallization and ITC experiments. Laëtitia Misson cloned all other constructs, purified all proteins for kinetic studies and performed enzyme kinetics and HPLC analysis. I designed and performed all protein crystallography experiments, performed data processing, structure solution, refinement, model building and analysis of the structures. I also performed the ITC measurements (during my master thesis). Florian P. Seebeck performed structural analysis and wrote the manuscript with the help of L. Misson and me. Wulf Blankenfeldt assisted in experiment design, structure solution of $\mathrm{EgtD_{wt}}$ apo and in writing the manuscript. F. P. Seebeck designed and managed the project.

ChemPubSoc Europe
DOI: 10.1002/cbic.201402522
CHEMBIOCHEM
Full Papers

Ergothioneine Biosynthetic Methyltransferase EgtD Reveals the Structural Basis of Aromatic Amino Acid Betaine Biosynthesis

Allegra Vit,[b, c] Laëtitia Misson,[a] Wulf Blankenfeldt,[b, c] and Florian P. Seebeck*[a]

Ergothioneine is an *N*-α-trimethyl-2-thiohistidine derivative that occurs in human, plant, fungal, and bacterial cells. Biosynthesis of this redox-active betaine starts with trimethylation of the α-amino group of histidine. The three consecutive methyl transfers are catalyzed by the S-adenosylmethionine-dependent methyltransferase EgtD. Three crystal structures of this enzyme in the absence and in the presence of *N*-α-dimethylhistidine and S-adenosylhomocysteine implicate a preorganized array of hydrophilic interactions as the determinants for substrate specificity and apparent processivity. We identified two active site mutations that change the substrate specificity of EgtD 10^7-fold and transform the histidine-methyltransferase into a proficient tryptophan-methyltransferase. Finally, a genomic search for EgtD homologues in fungal genomes revealed tyrosine and tryptophan trimethylation activity as a frequent trait in ascomycetous and basidomycetous fungi.

Introduction

Ergothioneine is a redox-active betaine of 2-thiohistidine occurring in bacteria, fungi, plants, and mammals (Scheme 1).[1] The precise physiological role of ergothioneine is not clear, but several observations indicate that this amino acid derivative might participate in cellular redox defense.[1c,2]

Mycobacteria such as the pathogenic *Mycobacterium tuberculosis* or the saprophytic *Mycobacterium smegmatis* produce ergothioneine in a five-step process starting with N-α-trimethylation of histidine catalyzed by the *S*-adenosylmethyltransferase (SAM) EgtD. The resulting histidine betaine is then subjected to oxidative sulfurization of the imidazole side chain at C2 by sequential action of the enzymes EgtB, EgtC, and EgtE (Scheme 1).

EgtD is the first known methyltransferase that converts a free amino acid directly into its corresponding betaine. In contrast, the common osmolyte and methyl donor glycine betaine is usually produced through a multi-enzyme pathway starting from the precursor choline (Scheme 1).[3] This pathway is specific for glycine betaine and is hardly adaptable to the production of betaines from other amino acids. Nevertheless, amino acid betaines constitute a diverse and ubiquitous class of natural products. The betaine of tryptophan (hypaphorine), for example, is a neuroactive component in peanuts and other Leguminosae species and consequently part of our daily nutrition.[4] In addition, hypaphorine is the most abundant soil alkaloid, because ectomycorrhizal fungi secrete hypaphorine as a suppressor for auxin (indol-3-acetic acid)-controlled root hair development by their host plants.[5] Isolation of hydroxylated or halogenated hypaphorine derivatives from marine invertebrates, or of trimethylated tyrosine or dihydroxyphenylalanine from fungal species, adds to the abundance and diversity of this class of natural products.[6] The biosynthetic origins of amino acid betaines other than trimethylglycine are unknown—until now.

In this report we show that the Methyltransf_33 protein family consists of synthases for histidine, tryptophan, and tyrosine betaine. We discuss the crystal structure and catalytic activity of EgtD in comparison with a double mutant of EgtD that displays specific and proficient tryptophan methyltransfer activity. On the basis of this analysis we identified tyrosine and tryptophan betaine synthases in ascomycetous and basidiomycetous fungi, thus paving the way for systematic and genome-based investigation of amino acid betaines in living systems.

Results and Discussion

EgtD catalyzes trimethylation of histidine with remarkable apparent processivity.[1b] An EgtD-catalyzed reaction involving one equivalent of histidine and one equivalent of *S*-adenosyl-L-methionine (SAM), for example, produces 88 ± 4 % trimethylhistidine, together with 13 ± 9 % dimethylhistidine, but less than 1 % *N*-α-methylhistidine. This processivity allows EgtD to route a fraction of the intracellular histidine pool towards ergothioneine biosynthesis, without interfering with ribosomal protein synthesis. The following analysis suggests that the observed

---

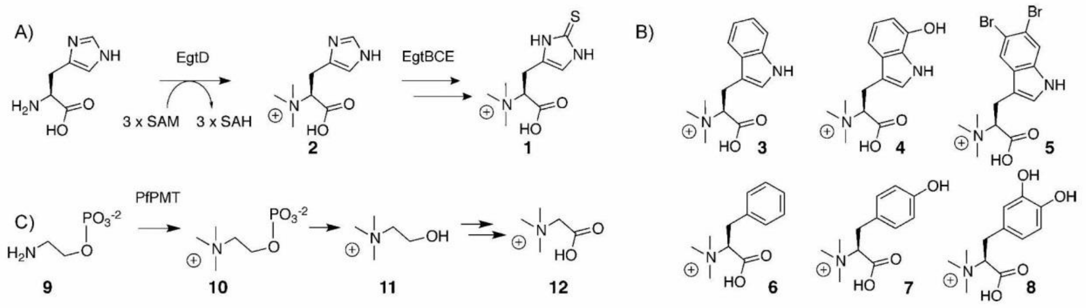
[a] *L. Misson,*[+] *Prof. Dr. F. P. Seebeck*
Department for Chemistry, University of Basel
St. Johanns-Ring 19, 4056, Basel (Switzerland)
E-mail: florian.seebeck@unibas.ch

[b] *A. Vit,*[+] *Prof. Dr. W. Blankenfeldt*
^{3}*Structure and Function of Proteins*
Helmholtz Centre for Infection Research
Inhoffenstrasse 7, 38124 Braunschweig (Germany)

[c] *A. Vit,*[+] *Prof. Dr. W. Blankenfeldt*
Previous address: Department of Biochemistry, University of Bayreuth
Universitätsstrasse 30, 95447 Bayreuth (Germany)

[+] *These authors contributed equally to this work.*

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201402522.



Scheme 1. A) The first and essential step in ergothioneine biosynthesis is catalyzed by EgtD, an S-adenosylmethionine-dependent methyltransferase.[1b] B) Examples of additional aromatic amino acid betaines isolated from fungi, plants, and marine invertebrates: hypaphorine (**3**),[5] hypaphorine derivatives **4** and **5**,[6c] phenylalanine betaine (**6**), tyrosine betaine (**7**),[6a] and sticticine (**8**).[6b] C) Biosynthesis of glycine betaine (**12**) by the malarial parasite *Plasmodium falciparum*: this synthesis starts with trimethylation of phosphatidylethanolamine (**9**) to phosphocholine (**10**),[3b] followed by dephosphorylation to choline (**11**) and two oxidation steps via the betaine aldehyde to glycine betaine.[3a]

Table 1. Kinetic parameters of the aromatic amino acid betaine synthases EgtD, Ybs, and EgtD$_{\mathrm{M252A,E282A}}$.[a]

	EgtD		EgtD$_{\mathrm{E285A}}$		Ybs		EgtD$_{\mathrm{M252A,E285A}}$	
Substrates	k_{cat} [s^{-1}]	k_{cat}/K_M [M^{-1} s^{-1}]	k_{cat} [s^{-1}]	k_{cat}/K_M [M^{-1} s^{-1}]	k_{cat} [s^{-1}]	k_{cat}/K_M [M^{-1} s^{-1}]	k_{cat} [s^{-1}]	k_{cat}/K_M [M^{-1} s^{-1}]
His	5.8×10^{-1}	5.3×10^{3}	1.0×10^{-1}	4.0×10^{1}	2.0×10^{-2}	7.9×10^{2}	–	1.9×10^{0}
N-α-His	2.3×10^{-1}	1.3×10^{4}	–	–	–	–	–	–
N,N-α-His	4.3×10^{-1}	1.7×10^{4}	–	–	–	–	–	–
Phe	–	2.4×10^{0}	–	–	1.4×10^{-1}	2.5×10^{1}	3.4×10^{-1}	1.0×10^{2}
Tyr	–	1.1×10^{0}	–	–	1.1×10^{-1}	6.3×10^{3}	–	1.7×10^{1}
Dopa	–	2.2×10^{0}	–	–	1.0×10^{-1}	6.6×10^{2}	–	4.5×10^{0}
Trp	–	2.0×10^{0}	3.0×10^{-2}	4.0×10^{2}	–	7.2×10^{0}	1.1×10^{-1}	5.5×10^{3}

[a] Data represent averages from multiple measurements. The standard error is less than 20% of the average value. Reaction conditions: 50 mM Tris•HCl pH 8.0, 26 °C, 200 μM SAM, 1 u S-adenosylhomocysteine nucleosidase, 1 u adenine deaminase. Ybs: tyrosine betaine synthase from *A. nidulans*.

processivity is a result of increased affinity for the methylated intermediates, rather than increased catalytic efficiency of the second and third methylation steps (Table 1).

We employed a UV-based assay[7] to monitor the rate of EgtD-catalyzed consumption of SAM in the presence of histidine, *N*-α-methyl- or *N*-α-dimethylhistidine. All three EgtD-catalyzed reactions proceed with similar turnover numbers (k_{cat}), and with only two- and threefold increased catalytic efficiency in the presence of *N*-α-methyl- and *N*-α-dimethylhistidine, respectively (Table 1). These slight differences among the three substrates could hardly explain the observed processivity. On the other hand, isothermal titration calorimetry (ITC) showed that EgtD is a four- and 70-times stronger binder of α-*N*-methylhistidine and *N*-α-dimethylhistidine, respectively, than of histidine (Table 2). A similar but weaker trend was observed in the presence of saturating concentrations of *S*-adenosyl-L-homocysteine (SAH).

Table 2. Isothermal calorimetry titration of EgtD.[a]

Titrant	K_D [μM]	ΔH [kcal mol^{-1}]	ΔS [cal mol^{-1} deg^{-1}]
His	290 ± 14	−8 ± 1	−11
N-α-methyl	70 ± 30	−13 ± 3	−25
N-α-dimethyl	4 ± 2	−5 ± 3	8
His[b]	37 ± 1	−10 ± 1	−13
N-α-methyl[b]	14 ± 7	−11 ± 2	−15
N-α-dimethyl[b]	2 ± 1	−27 ± 2	−64
SAH	210 ± 20	−12 ± 4	−25
SAM	270 ± 20	−19 ± 1	−49

[a] Reaction conditions: 20 mM Tris•HCl pH 7.5, 150 mM NaCl, 25 °C, 100 μM EgtD in cell, 5 mM ligand in syringe. [b] EgtD solution contained 7 mM *S*-adenosyl-L-homocysteine (SAH).

To investigate the structural basis for this substrate selectivity we determined the structure of EgtD as an apoprotein (resolution: 1.75 Å), as a binary complex with *N*-α-dimethylhistidine (resolution: 1.9 Å), and as a ternary complex with *N*-α-dimethylhistidine and SAH (resolution 1.5 Å, Tables S1 and S2 in the Supporting Information). The asymmetric unit cells of these crystals each contained two protein chains arranged either in the $P2_12_12_1$ space group (apo form, binary complex) or in the $P2_1$ space group (ternary complex; Figure 2, below). The overall structure consists of a Rossmann-fold domain typical for SAM-dependent methyltransferases.[8] The first 60 residues and a 100-residue insertion after β-strand 5 of the central β-sheet combine to form a second domain. The cleft between the two domains harbors the binding site for histidine and SAM.

Phosphoethanolamine methyltransferase (PfPMT) from *Plasmodium falciparum* and EgtD both catalyze SAM-dependent trimethylation of primary amines in the context of amino acid betaine biosynthesis (Scheme 1). The two structures share a su-

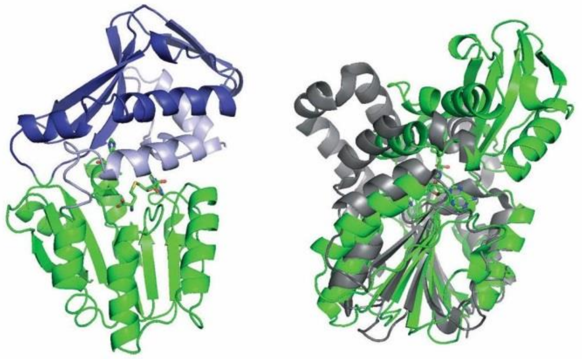


Figure 1. Left. The structure of EgtD in complex with SAH and *N*-α-dimethylhistidine. The substrate-binding site is located in a cleft between a Rossman-fold domain (green) and a domain formed by residues 15–60 (light blue) and 196–286 (dark blue). Right: Superimposition of EgtD (green) and phosphoethanolamine methyltransferase (PfPMT) [EC:2.1.1.103] from *Plasmodium falciparum* (gray).[3b]

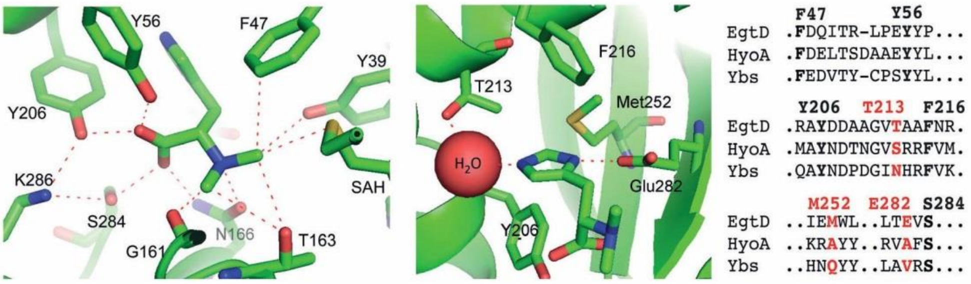
perimposable Rossman-fold domain and similar SAM-binding sites (Figure 1). The second domains, which define substrate specificity, are unrelated;[3b] this suggests that glycine and histidine betaine biosynthesis emerged through separate evolutionary channels.

Closer inspection of the active site in EgtD reveals a network of seven hydrogen bonds between the enzyme and the N-α-dimethylamino acid moiety of the substrate (Figure 2). The carboxylate group hydrogen-bonds to the side chains of Tyr206 (2.6 Å), Ser284 (2.8 Å), Tyr56 (2.6 Å), and Asn166 (2.9 Å). Asn166 also approaches the N-α-dimethylamino moiety of the substrate at hydrogen-bonding distance (2.8 Å), thus suggesting that the N-α-dimethylamino group is protonated.

The interactions between the active site and the methyl groups of the *N*-α-dimethylhistidine should give some indication as to why the methylated intermediate is a 70-times better EgtD ligand than histidine. In the ternary structure (Figure 2), one *N*-α-methyl group is in van der Waals contact with the sulfur atom of SAH (3.4 Å), the oxygen atoms in the side chains of Thr163 (3.5 Å) and Tyr39 (3.5 Å), and the phenyl ring of Phe47 (4.0 Å). The second *N*-α-methyl group makes close contact to the backbone carbonyl group of Gly161 (3.0 Å). An identical set of interactions is observed in the binary complex. This dense array of short contacts between nucleophilic side chains and the polarized *N*-α-methyl groups provides a viable explanation for preferential binding of *N*-α-methyl- and *N*-α-dimethylhistidine versus histidine.[9,11a,12]

The side chain of the substrate is clamped by two edge-to-face interactions to Phe216 and Tyr206 (Figure 2). Nτ of the imidazole ring connects to Thr213 through a water-mediated hydrogen bond, and Nπ shares a proton with Glu282 (2.5 Å). This last interaction is essential for efficient EgtD activity: mutation of Glu282 to alanine reduces EgtD activity 10^3-fold because of a 200-fold increase in K_M (EgtD$_{E282A}$, Table 1). Consistently this glutamate residue is highly conserved among all EgtD homologues that are co-encoded with EgtB and are therefore clearly involved in ergothioneine biosynthesis. However, BLAST searches for EgtD homologues in fungal genomes also revealed sequences with apolar residues at position 282. To us this was the first indication that distant EgtD homologues might have diversified to trimethylate amino acids other than histidine.

With the structure of EgtD—the first structure of any member of the Methyltransf_33 protein family—to hand we can now test this idea. EgtD homologues are present in in almost 700 sequenced species.[10] Bacterial genomes in this group usually each encode a single EgtD homologue; these are co-encoded with EgtB and share nearly identical sets of active site residues with EgtD from *M. smegmatis*. It is therefore safe to assume that these bacterial homologues are bona fide histidine methyltransferases involved in ergothioneine biosynthesis. Fungal genomes, on the other hand, encode on average two EgtD homologues. One homologue is usually encoded as a fusion protein with EgtB; this is consistent with the notion that ergothioneine biosynthesis is a highly conserved trait among fungi.[1b,11] The second EgtD homologues are characterized by considerable variation at positions 213, 252, and 282



```
      F47      Y56
EgtD  .FDQITR-LPEYYP...
HyoA  .FDELTSDAAEYYL...
Ybs   .FEDVTY-CPSYYL...

       Y206  T213  F216
EgtD  .RAYDDAAGVTAAFNR.
HyoA  .MAYNDTNGVSRRFVM.
Ybs   .QAYNDPDGINHRFVK.

        M252  E282 S284
EgtD  ..IEMWL..LTEVS...
HyoA  ..KRAYY..RVAFS...
Ybs   ..HNQYY..LAVRS...
```

Figure 2. Left: The amino acid moiety of *N*-α-dimethylhistidine is recognized through an array of seven hydrogen bonds, three of which are C–H···O bonds to the protonated *N*-α-dimethyl moiety. Middle: Two additional hydrogen bonds to Glu282 and a Thr213-coordinated water molecule immobilize the imidazole side chain of the substrate. Right: Abbreviated sequence alignment of the histidine methyltransferase from *M. smegmatis* (EgtD), a tyrosine methyltransferase from *Aspergillus nidulans* with 28 % sequence identity to EgtD$_{M. smegmatis}$ (Ybs, AN8594.2), and a putative tryptophan methyltransferase from *Dichomitus squalens* with 26 % sequence identity to EgtD$_{M. smegmatis}$ (HyoA, DICSQ_157002). In red: the active sites of the three enzymes differ most significantly at positions 213, 252 and 282.

(EgtD numbering, Figure 2), thus indicating that these methyltransferases might be specific for alternative substrates.

One such subclass contains the uncharacterized protein AN8594.2 from the ascomycetous fungus *Aspergillus nidulans* (Figure 3, below). This protein consists of a methyltransferase domain with 28% sequence identity to EgtD$_{M.\ smegmatis}$ and a putative iron(II)-dependent hydroxylase domain (TauD-like domain). We produced the methyltransferase domain of AN8594.2 in *E. coli* and assayed its methyltransferase activity for each of the 20 standard amino acids.

This screen identified tyrosine as the preferred substrate and tyrosine betaine as the corresponding product. The fungal enzyme catalyzed tyrosine methylation with a catalytic efficiency comparable to that of EgtD-catalyzed histidine methylation (Table 1). Phenylalanine, dihydroxyphenylalanine (DOPA), and histidine are ten times less efficient substrates, and tryptophan or non-aromatic amino acids are at least 10^3 times poorer substrates. A structural model of AN8594.2 based on the EgtD structure suggests that the fungal enzyme recognizes tyrosine through hydrogen bonds between the phenol side chain and the asparagine at position 213 (EgtD numbering, Figure 2) and/or a glutamine side chain at position 252.

On the basis of the observed in vitro activity we suggest that this fungal EgtD homologue is a tyrosine betaine synthase (Ybs). Presently, we have no information about the function of the TauD-like domain. However, we note that the betaines of tyrosine and dihydroxyphenylalanine (sticticine, **8**, Scheme 1) are both known metabolites from ascomycetous fungi.[6a,12] It therefore seems possible that the TauD-like domain is responsible for the transformation of tyrosine betaine into sticticine.

A BLAST search for EgtD homologues in genomes from basidiomycetous fungi revealed a different subclass of potential amino acid betaine synthases (Figure 3). One member of this subclass is the uncharacterized protein DICSQ_157002 from the wood-rotting fungus *Dichomitus squalens*. Sequence alignment with EgtD$_{M.\ smegmatis}$ indicates that this protein contains smaller residues at positions 213, 252, and 282 (EgtD numbering, Figure 2). The corresponding substrate-binding pocket should be more spacious than that in EgtD and might accommodate tryptophan as a substrate. On the basis of this alignment we surmised that DICSQ_157002 might be a hypaphorine synthase (HyoA). In support of this prediction we were able to detect hypaphorine as a metabolite in *D. squalens* growing on maltose extract agarose (Figure 4). The isolated compound co-eluted with authentic hypaphorine on reversed-phase HPLC, the elemental composition as determined by HRMS (ESI) was according to expectation, and the compound fragmented to indol-3-acrylic acid during MS analysis. Discovery of hypaphorine in a saprophytic fungus is surprising and suggests that this redox-active alkaloid might have a physiological role in addition to supporting fungus-to-plant communication in the ectomycorrhiza.[13]

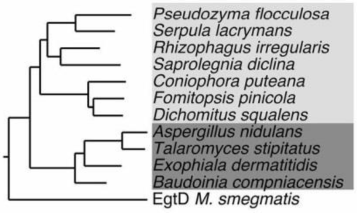


Figure 3. Basidomycetes (light gray) and Ascomycetes (gray) encode putative EgtD homologues with tryptophan (light gray) and tyrosine (gray) specificity. *Serpula lacrymans, Saprolegnia diclina, Coniophora puteana, Fomitopsis pinicola*, and *Dichomitus squalens* are saprophytic fungi, whereas *Rhizophagus irregularis* is a mycorrhizal fungus.

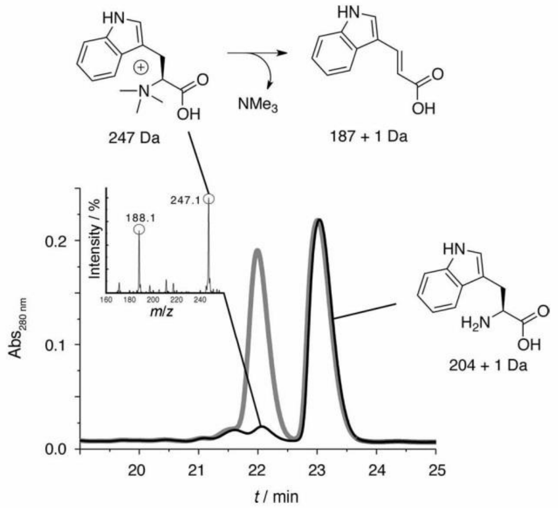


Figure 4. Detection of hypaphorine in malt-extract-grown *D. squalens* by ESI-MS and RP-HPLC (black line). Hypaphorine was identified by ESI-MS (*m/z* calcd 247.14; measured 247.13), by HRMS (*m/z* calcd 247.1441; measured 247.1441, consistent with an elemental composition $C_{14}H_{19}N_2O_2^+$), by co-elution with an authentic sample of hypaphorine (gray line), and by ESI-induced fragmentation to trimethylamine and indol-3-acrylic acid (*m/z* calcd 188.1; measured 188.1).

Because we were unable to produce HyoA from *D. squalens* in *E. coli* we engineered an EgtD variant with an active site configuration similar to that in the fungal protein (EgtD$_{\text{M252A,E282A}}$, Figure 2). The two mutations Met252 to Val and Glu282 to Ala increased tryptophan-specific activity more than 2×10^3-fold and reduced the histidine-specific activity 3×10^3-fold. Despite this dramatic change in substrate specificity EgtD$_{\text{M252A,E282A}}$ still discriminates against other possible substrates such as phenylalanine, tyrosine, or DOPA (Table 1) with strictness similar to those of the evolved enzymes EgtD and Ybs.

To explain the structural basis for this excellent substrate specificity we also determined the crystal structure of EgtD$_{\text{M252A,E282A}}$ in complex with tryptophan and SAH (resolution 1.8 Å, Figure 5). The complex structure contains weak but significant residual electron density due to tryptophan bound in the mutant active site (Figure S1). The binding mode is very similar to that described for the complex between EgtD and *N*-α-dimethylhistidine: the indole ring is pinched by the aromatic rings of Tyr206 and Phe216 (Figure 2) and engages in a water-mediated hydrogen bond to Thr213 (Figure 5). This combina-

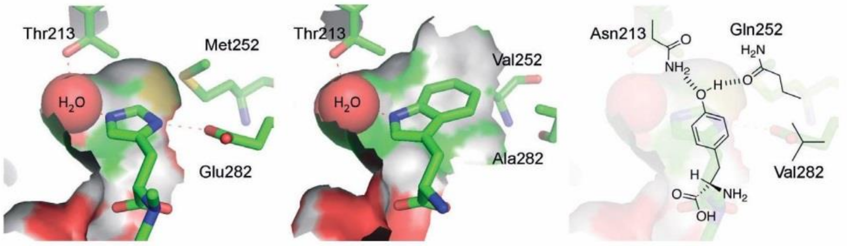


Figure 5. Comparison of substrate binding sites of EgtD wild type in complex with *N*-α-dimethylhistidine (left) and of EgtD$_{M252,E282A}$ in complex with tryptophan (middle). A model of the tyrosine methyltransferase from *A. nidulans* (right) suggests that tyrosine is recognized by hydrogen bonding to the residues Asn213 and Gln252 (residue numbering based on the EgtD sequence).

tion of hydrophobic and hydrophilic interactions provides the basis for the observed tryptophan specificity of EgtD$_{M252A,E282A}$. Tyrosine or phenylalanine could not establish a hydrogen bond to Thr213, and are therefore rejected as substrates.

The almost 10^7-fold change in enzyme specificity ($[k_{cat,His}/K_{M,His}]/[k_{cat,Trp}/K_{M,Trp}]$) as a result of two point mutations illustrates the ease with which new enzyme activity can emerge.[14] More specifically, this observation suggests that the Methyltransf_33 protein family might include many more amino acid betaine synthases.

Conclusion

We have described the crystal structure of EgtD, thus providing a first view into the ubiquitous Methyltransf_33 protein family. The structure in complex with the ligands *N*-α-dimethylhistidine and SAH revealed determinants for substrate selectivity and product specificity. On the basis of this analysis we identified tyrosine betaine synthases and tryptophan betaine synthases in fungal genomes; this suggests that aromatic amino acid betaines might be common metabolites in eukaryotic microorganisms. Finally, EgtD is an essential enzyme for ergothioneine biosynthesis in Mycobacteria.[2c] Inhibition of this pathway might present a target for new therapeutics against *Mycobacterium tuberculosis*. This structure provides the basis for rational design of ergothioneine biosynthesis inhibitors to test this concept.

Experimental Section

Protein production, crystallization, data collection, structure solution, and refinement: Native and seleno-L-methionine-labeled EgtD, as well as EgtD$_{M252A,E282A}$, were produced as described previously.[15] Crystallization of native and seleno-L-methionine-labeled EgtD was achieved as described previously.[15] For co-crystallization of native EgtD and EgtD$_{M252A,E282A}$ with substrates, new crystallization conditions had to be determined, and screens were set up with a Phoenix nano-dispensing robot (Art Robbins Instruments). Crystals of EgtD with *N*-α-dimethylhistidine (EgtD_AVI) were obtained by mixing protein and ligand at concentrations of 5 mM *N*-α-dimethylhistidine and 20 mg mL^{-1} EgtD in a drop (0.5 µL) mixed 1:1 with reservoir [sodium phosphate (1.6 M), dipotassium phosphate (0.4 M), sodium phosphate citrate (0.1 M), pH 4.2]. Crystals of EgtD with *N*-α-dimethylhistidine and SAH (EgtD_AVI_SAH) were obtained by mixing both compounds with native EgtD to a final concentration of 5 mM *N*-α-dimethylhistidine, 5 mM SAH, and 20 mg mL^{-1} EgtD. This solution was mixed 1:1 with reservoir [magnesium chloride (0.2 M), Tris·HCl (pH 8.5, 0.1 M), PEG 4000 (30%, *w*/*v*)] and equilibrated against reservoir (500 µL). Crystals of EgtD$_{M252A,E282A}$ with L-tryptophan and SAH (EgtD$_{M252A,E282A}$_TRP_SAH) were obtained by preincubation of EgtD$_{M252A,E282A}$ (15 mg mL^{-1}) with L-tryptophan (5 mM) and SAH (5 mM). This solution was mixed 1:1 with reservoir [PEG 3350 (20%, *w*/*v*) and magnesium acetate (0.2 M)]. All crystallization experiments were performed at room temperature. Prior to data collection, crystals of EgtD_AVI were cryoprotected in reservoir solution supplemented with glycerol (5%, *v*/*v*). No cryoprotection was required for crystals of EgtD_AVI_SAH. Crystals of EgtD$_{M252A,E282A}$_TRP_SAH were cryoprotected in reservoir solution supplemented with PEG 3350 (35%, *w*/*v*).

Diffraction data for seleno-L-methionine-labeled and native apo form crystals of EgtD were collected at beamline PXII of the Swiss Light Source (SLS at the Paul Scherrer Institute, Villingen, Switzerland) as described previously.[15] Diffraction data for EgtD_AVI, EgtD_AVI_SAH, and EgtD$_{M252A,E282A}$_TRP_SAH were collected at 100 K at a wavelength of $\lambda = 0.918$ Å on beamline BL 14.1 of the BESSY II synchrotron (Helmholtz-Zentrum Berlin, Germany).

The single anomalous dispersion (SAD) data for the seleno-L-methionine-labeled protein were processed as described[15] and led, with the aid of autoSHARP,[16] to a structure solution of the EgtD apo form. The model was then automatically built with ARP/wARP 7.0.[17]

Diffraction data for the EgtD complex structures (EgtD_AVI, EgtD_AVI_SAH, EgtD$_{M252A,E282A}$ _TRP_SAH) were indexed with XDS[18] and then processed with COMBAT and SCALA from the CCP4 program suite.[19] In the case of EgtD_AVI, structure solution was carried out by rigid body refinement in REFMAC5[19b] with use of the EgtD apo structure. The apo structure of EgtD was also used in MOLREP[20] to solve the structure of EgtD_AVI_SAH. This structure was used to phase the structure of EgtD$_{M252A,E282A}$ by rigid body refinement in REFMAC5.[19b]

For structure refinement, ligands were prepared with ProDRG[21] for REFMAC5[19b] and eLBOW[22] for phenix.refine.[23] Model building was done in COOT.[23] After refinement and model building, all structures were checked with MolProbity.[24] All figures were prepared with PyMOL.[25] Details of the data collection and refinement statistics can be found in Tables S1 and S2. Coordinates and structures factors have been deposited in the protein data bank with IDs 4PIM, 4PIN, 4PIO, and 4PIP.[26]

Isothermal titration calorimetry (ITC): The different compounds used for ITC measurements were dissolved in Tris/HCl (pH 7.5, 20 mM), NaCl (150 mM). Directly before the measurements, EgtD was eluted from a NAP5 Sephadex column (GE Healthcare) primed with the exact same buffer to enhance baseline stability. Protein solutions were then diluted to a concentration of 100 µM and transferred into the sample cell of an isothermal titration calorimeter (VP200-ITC system, MicroCal LLC). Compounds (L-histidine, *N*-methyl-L-histidine, *N*-α-dimethylhistidine, SAM, SAH) were used in different concentrations (5, 7, or 10 mM) and were added to the protein solution by syringe/stirrer in 2 μL steps every 2 min. All ITC measurements were conducted at 25 °C. Data were plotted as the power needed to maintain the reference and sample cell at the same temperature against time and as kcal mol^{-1} of injectant against the molar ratio of ligand and protein. The program Origin7 (OriginLab Corporation) was used to analyze the data.

Cloning and production of Ybs: The gene coding for the methyltransferase domain (Methyltransf_33) of the protein AN8594.2 from *Aspergillus nidulans* was purchased from GeneScript with codon adaptions for optimal production in *Escherichia coli*. This fragment was cloned into a pOPIN-expression vector by using restriction-free protocols. The resulting vector encodes Ybs as a fusion with an N-terminal hexahistidine tag. Ybs was produced in *E. coli* Bl21 cell by the same procedure as described for EgtD.[15]

Construction of EgtD variants—EgtD$_{\text{E282A}}$ and EgtD$_{\text{M252A,E282A}}$: The two variants were constructed by using mutagenesis primers: M252Vs: 5′-ATC GAG GTT TGG TTG CGT GCC CGC A-3′, M252Va: 5′-CAC GCA ACC AAA CCT CGA TGC GTT CCT-3′, E282As: 5′-GAT GCT CAC CGC AGT GTC CTG CAA GTT-3′, E282Aa: 5′-TGC AGG ACA CTG CGG TGA GCA TCT CCT-3′, EgtDs: 5′-ATA TCA TAT GGC GCT CTC ACT GGC CAA-3′, EgtDa: 5′-ATA TCT CGA GTC ACC GCA CCG CCA GCG ACA-3′. The gel-purified fragment was digested with NdeI and XhoI restriction enzymes and ligated to a modified pET19 expression vector. The proteins were produced and purified by the same protocols as used for the EgtD wild type.

Methyltransferase assay: Methyltransferase activity was determined by published protocols.[7] Reactions were monitored in a 2 mm quartz cuvette at 25 °C and 265 nm. The 200 μL reaction mixtures each contained Tris·HCl (pH 8, 50 mM), NaCl (50 mM), $MnBr_2$ (100 μM), SAM (200 μM), adenine deaminase (4.5 μM), AdoHcy nucleosidase (12 μM), methyltransferase (2–4 μM), and the appropriate amino acid at a concentration of 5–500 μM. The data were fitted to the function $v = k_{\text{cat}}[\text{s}]/(K_{\text{M}} + [\text{s}])$. Methyltransferase products were identified by ESI-MS (trimethylhistidine: *m/z* calcd: 198.12; found: 198.1; trimethyltyrosine: *m/z* calcd: 224.13; found: 224.1; trimethyltryptophan: *m/z* calcd: 247.14; found: 247.1; trimethylated dihydroxyphenylalanine: *m/z* calcd: 240.12; found: 240.1).

Adenosylhomocysteine nucleosidase: AdoHcy was produced in *E. coli* BL21 cells grown in LB medium [kanamycin ($50\ \text{mg L}^{-1}$), chloramphenicol ($34\ \text{mg L}^{-1}$)] and induced with isopropyl-β-D-thiogalactopyranoside (IPTG; 1 mM) at 37 °C for 3 h. Cells were pelleted and resuspended in phosphate buffer (pH 8.0, 50 mM) and NaCl (300 mM). Cells were lysed at 4 °C by sonication, and the cleared lysate was mixed with Ni^{II}-NTA agarose (Qiagen) at 4 °C for 20 min. The agarose beads were washed with phosphate buffer containing imidazole (10 and 20 mM). The protein was eluted in an imidazole solution (250 mM). The purified eluted protein was dialyzed into Tris·HCl buffer (pH 8.0, 50 mM), NaCl (50 mM), β-mercaptoethanol (5 mM) and stored at −80 °C. To approximate the concentration of the prepared protein, we used a calculated molar absorption coefficient of $\varepsilon_{280}(\text{AdoNuc}) = 5960\ \text{M}^{-1}\,\text{cm}^{-1}$.

Adenine deaminase: Adenine deaminase was produced in *E. coli* BL21 cells grown in LB medium [kanamycin ($50\ \text{mg L}^{-1}$), chloramphenicol ($34\ \text{mg L}^{-1}$)] with bipyridine (50 mM) and induced with IPTG (1 mM) and $MnCl_2$ (1 mM) at 37 °C for 3 h. Cells were pelleted and resuspended in phosphate buffer (pH 8.0, 50 mM), Tween 20 (0.1 %), and NaCl (300 mM). Cells were lysed at 4 °C by sonication, and the cleared lysate was mixed with Ni^{II}-NTA agarose (Qiagen) at 4 °C for 20 min. The agarose beads were washed with phosphate buffer containing imidazole (10 and 20 mM), glycerol (10 %), and Tween 20 (0.1 %). The protein was eluted in a solution of imidazole (250 mM), glycerol (10 %), and Tween 20 (0.1 %). The purified eluted protein was dialyzed into Tris·HCl buffer (pH 8.0, 50 mM), NaCl (50 mM), β-mercaptoethanol (5 mM) and stored at −80 °C ($\varepsilon_{280}(\text{deaminase}) = 41\,370\ \text{M}^{-1}\,\text{cm}^{-1}$).

Detection of hypaphorine production: *D. squalens* was purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSM9615) as an actively growing culture. The fungus was grown on malt extract agar at 22 °C for three weeks. Mycelium (13 g) was extracted with a $MeOH/H_2O$ (9:1) mixture. The concentrated extract was analyzed by HPLC (RP C_{18}).

Note Added in Proof

Jeong et al. published the crystal structure of EgtD in the apo form, in complex with histidine or histidine and SAH (J.-H. Jeong, H. J. Cha, S.-C. Ha, C. Rojviriya, Y.-G. Kim, *Biochem. Biophys. Res. Commun.* **2014**, *452*, 1098–1103, DOI: 10.1016/j.bbrc.2014.09.058). These structures provide a valuable complement to the EgtD complex with SAH and *N*-α-dimethylhistidine described here.

Acknowledgements

The authors are indebted to Roger Goody, who enabled the initiation of this project at the MPI for Molecular Physiology in Dortmund. The authors would like to thank the X-ray communities at the MPI in Heidelberg and Dortmund and the University of Bayreuth for help with data collection, and the Swiss Light Source (PSI, Villingen) and BESSY II (Helmholtz Centre Berlin, Berlin) for beamline access. This project was supported by the Swiss National Foundation. F.P.S. is supported by the "Professur für Molekulare Bionik". A.V. is supported by the HZI Graduate School for Infection Research.

Keywords: amino acids · betaines · biosynthesis · ergothioneine · hypaphorine · methyltransferases

[1] a) D. B. Melville, M. L. Ludwig, E. Inamine, J. R. Rachele, *J. Biol. Chem.* **1959**, *234*, 1195–1198; b) F. P. Seebeck, *J. Am. Chem. Soc.* **2010**, *132*, 6632–6633; c) M. H. Bello, V. Barrera-Perez, D. Morin, L. Epstein, *Fungal Genet. Biol.* **2012**, *49*, 160–172; d) C. Pfeiffer, B. Surek, E. Schöming, D. Gründemann, *Food Chem.* **2011**, *129*, 1766–1769; e) C. Tanret, *C. R. Acad. Sci.* **1909**, *149*, 222–224; f) D. Gründemann, *Prev. Med.* **2012**, *54*, S71–S74; g) I. K. Cheah, B. Halliwell, *Biochim. Biophys. Acta* **2012**, *1822*, 784–793; h) R. C. Fahey, *Biochim. Biophys. Acta* **2013**, *1830*, 3182–3198.

[2] a) I. K. Cheah, R. L. Ong, J. Gruber, T. S. Yew, L. F. Ng, C. B. Chen, B. Halliwell, *Free Radical Res.* **2013**, *47*, 1036–1045; b) B. D. Paul, S. H. Snyder, *Cell Death Differ.* **2009**, *16*, S149–S149; c) C. Sao Emani, M. J. Williams, I. J. Wiid, N. F. Hiten, A. J. Viljoen, R.-D. D. Pietersen, P. D. van Helden, B. Baker, *Antimicrob. Agents Chemother.* **2013**, *57*, 3202–3207.

[3] a) T. H. Chen, N. Murata, *Plant Cell Environ.* **2011**, *34*, 1–20; b) S. G. Lee, Y. Kim, T. D. Alpert, A. Nagata, J. M. Jez, *J. Biol. Chem.* **2012**, *287*, 1426–1434.

[4] a) A. Tsopmo, A. D. Muir, *J. Agric. Food Chem.* **2010**, *58*, 8715–8721; b) B. O. Keller, B. T. Wu, S. S. Li, V. Monga, S. M. Innis, *J. Agric. Food Chem.* **2013**, *61*, 7654–7660.

[5] T. Kawano, *Plant Cell Rep.* **2003**, *21*, 829–837.

[6] a) T. Bernard, G. Goas, J. Hamelin, M. Joucla, *Phytochemistry* **1981**, *20*, 2325–2326; b) T. Bernard, M. Joucla, G. Goas, J. Hamelin, *Phytochemistry*

1980, *19*, 1967 – 1969; c) A. Mollica, M. Locatelli, A. Stefanucci, F. Pinnen, *Molecules* **2012**, *17*, 6083 – 6099.

[7] K. M. Dorgan, W. L. Wooderchak, D. P. Wynn, E. L. Karschner, J. F. Alfaro, Y. Q. Cui, Z. S. Zhou, J. M. Hevel, *Anal. Biochem.* **2006**, *350*, 249 – 255.

[8] R. Gana, S. Rao, H. Huang, C. Wu, S. Vasudevan, *BMC Struct. Biol.* **2013**, *13*, 6.

[9] a) S. Horowitz, L. M. A. Dirk, J. D. Yesselman, J. S. Nimtz, U. Adhikari, R. A. Mehl, S. Schreiner, R. L. Houtz, H. M. Al-Hashimi, R. C. Trievel, *J. Am. Chem. Soc.* **2013**, *135*, 15536 – 15548; b) S. Horowitz, R. C. Trievel, *J. Biol. Chem.* **2012**, *287*, 41576 – 41582.

[10] R. D. Finn, A. M. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heder, K. Hetherington, J. Mistry, E. L. Sonnhammer, J. Tate, M. Punta, *Nucleic Acids Res.* **2014**, *42*, D222 – D230.

[11] G. W. Jones, S. Doyle, D. A. Fitzpatrick, *Gene* **2014**, *549*, 161 – 170.

[12] C. A. Carollo, A. L. Calil, L. A. Schiave, T. Guaratini, D. W. Roberts, N. P. Lopes, G. U. Braga, *Fungal Genet. Biol.* **2010**, *114*, 473 – 480.

[13] C. J. Gagliardi, R. A. Binstead, H. H. Thorp, T. J. Meyer, *J. Am. Chem. Soc.* **2011**, *133*, 19594 – 19597.

[14] a) M. D. Toscano, K. J. Woycechowsky, D. Hilvert, *Angew. Chem. Int. Ed.* **2007**, *46*, 3212 – 3236; *Angew. Chem.* **2007**, *119*, 3274 – 3300; b) H. W. Wilks, K. W. Hart, R. Feeney, C. R. Dunn, H. Muirhead, W. N. Chia, D. A. Barstow, T. Atkinson, A. R. Clarke, J. J. Holbrook, *Science* **1988**, *242*, 1541 – 1544.

[15] A. Vit, L. E. Misson, W. Blankenfeldt, F. P. Seebeck, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2014**, *70*, 676 – 680.

[16] C. Vonrhein, E. Blanc, P. Roversi, G. Bricogne, *Methods Mol. Biol.* **2007**, *364*, 215 – 230.

[17] G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, *Nat. Protoc.* **2008**, *3*, 1171 – 1179.

[18] W. Kabsch, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125 – 132.

[19] a) P. R. Evans, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 282 – 292; b) G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 355 – 367.

[20] A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022 – 1025.

[21] A. W. Schüttelkopf, D. M. F. van Aalten, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 1355 – 1363.

[22] N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, *65*, 1074 – 1080.

[23] P. D. Adams, P. V. Afonie, G. Bunkoczi, V. B. Chen, I. W. Davies, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwillinger, P. H. Zwart, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 213 – 221.

[24] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 12 – 21.

[25] W. L. DeLano, *The PyMol User's Manual*, DeLano Scientific, San Carlos, CA, USA; **2002**, 2002.

[26] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235 – 242.

Received: September 9, 2014

12.2 Structure of the Sulfoxide Synthase EgtB from the Ergothioneine Biosynthetic Pathway

Publication B

Kristina V. Goncharenko*, Allegra Vit*, Wulf Blankenfeldt§ and Florian P. Seebeck§. Structure of the Sulfoxide Synthase EgtB from the Ergothioneine Biosynthetic Pathway. *Angew. Chem. Int. Ed. Engl*. **2015** Feb 23; 54(9): 2821-4.

*Both authors contributed equally to this work.
§Corresponding authors.

Author Contribution

Kristina V. Goncharenko performed molecular cloning, protein purification and enzyme kinetics of all proteins in this study. I designed and performed protein crystallographic experiments as well as structure solution, refinement and analysis. The manuscript was written by Florian P. Seebeck who was assisted by K. V. Goncharenko and me. Wulf Blankenfeldt helped in crystallographic analysis and manuscript writing. F. P. Seebeck designed and managed the project.

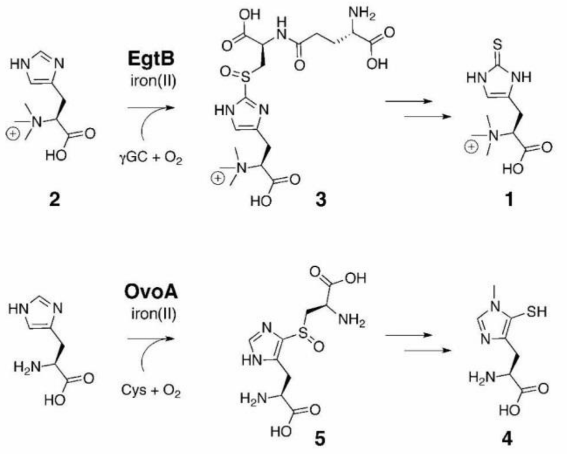
DOI: 10.1002/anie.201410045

Enzymatic C–S Coupling

Structure of the Sulfoxide Synthase EgtB from the Ergothioneine Biosynthetic Pathway**

Kristina V. Goncharenko, Allegra Vit, Wulf Blankenfeldt, and Florian P. Seebeck**

Abstract: *The non-heme iron enzyme EgtB catalyzes O_2-dependent C–S bond formation between γ-glutamyl cysteine and N-α-trimethyl histidine as the central step in ergothioneine biosynthesis. Both, the catalytic activity and the architecture of EgtB are distinct from known sulfur transferases or thiol dioxygenases. The crystal structure of EgtB from Mycobacterium thermoresistibile in complex with γ-glutamyl cysteine and N-α-trimethyl histidine reveals that the two substrates and three histidine residues serve as ligands in an octahedral iron binding site. This active site geometry is consistent with a catalytic mechanism in which C–S bond formation is initiated by an iron(III)-complexed thiyl radical attacking the imidazole ring of N-α-trimethyl histidine.*



Scheme 1. EgtB- and OvoA-catalyzed C–S bond formation and sulfoxidation between γ-glutamyl cysteine and N-α-trimethyl histidine (TMH) or between cysteine and histidine as the central steps in the syntheses of ergothioneine (**1**) and ovothiol (**4**), respectively.

Ergothioneine (**1**, Scheme 1) occurs in a broad range of prokaryotic and eukaryotic organisms, including humans and human pathogens such as *Mycobacterium tuberculosis*.[1] Higher eukaryotes absorb ergothioneine as a micronutrient, which stems from microbial production. The precise cellular function of ergothioneine is not known, but recent observations from animal, fungal, and bacterial cells suggest that this sulfur compound may be a protectant against oxidative stress.[2] *Mycobacteria* biosynthesize ergothioneine from glutamate, cysteine, and histidine.[1d,3] The central step in this pathway is catalyzed by the non-heme iron enzyme EgtB, which forms and sulfoxidizes a C–S bond between *N*-α-trimethyl histidine (**2**, TMH) and γ-glutamyl cysteine (γGC) with O_2 as the oxidant. Together with the ovothiol (**4**) biosynthesis enzyme OvoA (Scheme 1),[4] EgtB represents a distinct enzyme class (sulfoxide synthases) with no relation to sulfur oxidizing or C–S bond-forming iron enzymes such as cysteine dioxygenase or isopenicillin synthase.[5] Instead, sulfoxide synthases present a new entry to a rich collection of C–S bond-forming catalysts.[6]

To elucidate the structural basis for sulfoxide synthase activity, we determined the crystal structure of EgtB from *Mycobacterium thermoresistibile* ($\mathrm{EgtB}_{\mathrm{thermo}}$) in complex with iron and TMH, and as a quaternary complex with manganese, N-α-dimethyl histidine (DMH), and γGC. We produced this enzyme in *Escherichia coli* by following the protocols established for the EgtB homologue from *Mycobacterium smegmatis* ($\mathrm{EgtB}_{\mathrm{smegmatis}}$).[1d] The two enzymes share 81% sequence homology, both are monomeric (Figure S1 in the Supporting Information), and they display similar in vitro activities (Table 1). Both recombinant enzymes purified to a significant extent as iron-bound holoenzymes, as inferred from a ferrozine-based colorimetric assay ($\mathrm{EgtB}_{\mathrm{thermo}} > 95\%$, $\mathrm{EgtB}_{smegmatis} > 50\%$; Table S1 in the Supporting Information) and from titration of EgtB activity with $FeSO_4$ (Figure S2). In vitro activity was assayed in HEPES-buffered solutions in the presence of TMH, γGC, $FeSO_4$ (4 μM), ascorbate (2 mM), NaCl (100 mM), and TCEP (2 mM) at 25 °C. Formation of the sulfoxide product (**3**; Scheme 1 and Figure S3) was monitored by cation-exchange HPLC.[1d] Under these conditions, $\mathrm{EgtB}_{\mathrm{smegmatis}}$ and $\mathrm{EgtB}_{\mathrm{thermo}}$ catalyzed up to one turnover per second and remained active for hundreds of turnovers (Table 1 and Figure S3). In the absence of ascorbate, the activity of $\mathrm{EgtB}_{\mathrm{thermo}}$ ceased after 120 ± 20 turnovers, but fully recovered after the addition of 2 mM ascorbate (Figure S4).

---

[*] K. V. Goncharenko,[+] F. P. Seebeck
Department for Chemistry, University of Basel
St. Johanns-Ring 19, 4056 Basel (Switzerland)
E-mail: florian.seebeck@unibas.ch
A. Vit,[+] W. Blankenfeldt
Structure and Function of Proteins
Helmholtz Centre for Infection Research
Inhoffenstr. 7, 38124 Braunschweig (Germany)
E-mail: wulf.blankenfeldt@helmholtz-hzi.de

[+] These authors contributed equally to this work.

[**] We would like to thank the Swiss Light Source (PSI, Villingen) and BESSY II Synchrotron (Berlin (Germany)) for beamline access and the European research council and the HZI Graduate School for Infection Research for financial support. F.P.S. is supported by the "Professur für Molekulare Bionik" and by an ERC starting grant.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201410045.

Table 1: Kinetic parameters for the EgtB variants.[a]

Enzyme	Donor	Acceptor	$k_{cat,donor}$ [s^{-1}][b]	$k_{cat,donor}/K_M$ [$M^{-1}s^{-1}$]	$k_{cat,acceptor}$ [s^{-1}]	$k_{cat,acceptor}/K_M$ [$M^{-1}s^{-1}$]
$EgtB_{smegmatis}$	γGC	TMH	1.1	1.4×10^{4}	1.2	2.8×10^{4}
$EgtB_{thermo}$	γGC	TMH	8.6×10^{-1}	2.0×10^{4}	8.7×10^{-1}	2.2×10^{4}
$EgtB_{thermo}$	NGC	TMH	2.5×10^{-1}	2.3×10^{2}	n.a.	n.a.
$EgtB_{D416N}$	γGC	TMH	2.7×10^{-1}	1.3×10^{2}	1.1×10^{-1}	1.9×10^{4}
$EgtB_{D416N}$	NGC	TMH	1.0×10^{-1}	1.1×10^{3}	n.a.	n.a.

[a] γGC = gamma-glutamyl cysteine; TMH = N-α-trimethyl histidine; NGC = N-glutaryl cysteine; displayed values are averages from three independent measurements, errors correspond to less than 15% of the average value. [b] k_{cat} and k_{cat}/K_M were determined in the presence of cosubstrate at a concentration at least 3-fold higher than the corresponding K_M and in air-saturated buffers. Michaelis–Menten plots are displayed in the Supporting Information.

Incubation with γGC alone also inactivated $EgtB_{thermo}$ (Figure S4) and induced an absorption band at 565 nm (ε_{565} = 450 $M^{-1}cm^{-1}$; Figure S5), which also vanished after the addition of ascorbate. A similar absorption feature has been observed in cysteine-bound ferric cysteine dioxygenase, where it was assigned as a sulfur-to-iron(III) charge-transfer transition.[7] By contrast, no change in activity or absorbance was induced by incubation of $EgtB_{thermo}$ with TMH or buffer alone. We suspect that a γGC-dependent side reaction oxidizes EgtB to the inactive ferric state. The absorption at 565 nm of inactive $EgtB_{thermo}$ in complex with γGC provides a first indication that the sulfur atom of this ligand may directly interact with the catalytic iron center. The following structural analysis provides strong support for this notion.

We crystallized $EgtB_{thermo}$ as the apo protein (mthEgtB_apo), in complex with iron and TMH (mthEgtB_TMH), and in complex with manganese, DMH, and γGC (mthEgtB_DMH_γGC). The crystals belong to space group $P4_32_12$ with cell constants $a,b = 135$ and $c = 141$ Å. The apo structure was phased using the anomalous signal of the intrinsic sulfur atoms and the coordinated iron. The phases were later combined with native data to yield the structure of $EgtB_{thermo}$ at a resolution of 1.7 Å (apo). Ligand complexes of $EgtB_{thermo}$ were obtained to resolutions of 1.6 Å (TMH) and 1.98 Å (γGC and DMH). In all of the structures, the asymmetric unit includes two protein chains sharing only a relatively small and asymmetric interface.

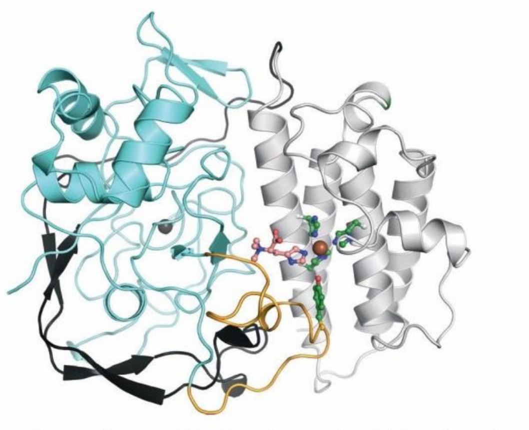


Figure 1. Cartoon diagram of $EgtB_{thermo}$ in complex with iron (brown) and TMH (pink). The protein consists of an N-terminal DinB domain (grey, residues 1–150), a two-stranded β-sheet region (black, residues 151–210), and a C-terminal C-type lectin domain (cyan/orange, residues 211–434). The active site, which contains a three-histidine facial triade (green; His51, His134, and His138) is formed between the DinB domain and residues 370–425 (orange)

The electron densities of the three $EgtB_{thermo}$ structures reveal a continuous polypeptide chain from Pro7 to Asp434 (Figure 1). Residues 7–150 fold to a DinB-like four-helix bundle with long linkers between helices 1 and 2 (18 residues), 2 and 3 (34 residues), and 3 and 4 (7 residues). The fourth helix is followed by an extended two-stranded β-sheet (residues 151–210) wrapped around the C-terminal domain, which adopts a C-type lectin fold (CLec).[8,9] This fold contains few secondary structure elements but instead is stabilized by a dense array of buried ionic interactions, such as salt bridges between Arg and Glu residues (Arg409:Glu196, Arg413:Glu296, Arg397:Glu300, and Arg428:Glu360). Furthermore, a calcium cation in the center of the C-terminal domain immobilizes six oxygen ligands from side chains and backbone amides (Met354 2.8 Å; Gly399 2.9 Å; Val358 2.6 Å; Gly356 2.7 Å; Gln353 3.6 Å; and Glu360 3.3 Å). This unusual loop-rich structure is conserved in at least two single-domain proteins with entirely different functions, with less than 30% sequence homology, and without dependence on transition metals.[8b,9] One of these structural homologues (r.m.s.d. of 0.741, Figure S6) is the human formylglycine-generating enzyme (FGE), which catalyzes the O_2-dependent posttranslational maturation of sulfatases.[8b] The second homologue is the diversity-generating retroelement variable protein TvpA from *Treponema denticola*.[9] Iron-dependent sulfoxide synthases such as EgtB and OvoA must have emerged from these FGE-like proteins through fusion with an N-terminal DinB domain. It is thus not surprising that the EgtB active site maps to the interface between the two domains.

The active site of $EgtB_{thermo}$ is housed in a 15 Å deep and 10 Å wide tunnel lined by residues 375 through 425 from the CLec domain and residues from helices 2 and 4 and the loops between helices 1, 2, and 3. At the bottom of this tunnel, three histidine residues from the DinB domain (residues 51, 134, and 138) coordinate the catalytic iron cation (Fe–N: 2.1 Å, 2.1 Å, and 2.3 Å). This structure revises our earlier prediction that Glu140 rather than His51 is the metal ligand.[1d,4a]

In hindsight, we note that this facial triad is fairly conserved among proteins from the DinB protein superfamily.[10] Several single-domain DinB proteins have been characterized as zinc-dependent thiol S-transferases or maleylpyruvate isomerases,[11] thus indicating that metal-catalyzed C–S bond formation may be a common activity among DinB-like proteins.[11a] However, the mechanisms of iron- and O_2-dependent oxidative C–S bond formation as catalyzed by EgtB and OvoA are most likely very different from that of zinc-promoted alkyl transfer.[6a,12] With this hypothesis in mind, we proceeded to study the substrate binding mode of $EgtB_{thermo}$.

In the ternary complex containing $EgtB_{thermo}$, TMH, and iron (ternary complex), the substrate imidazole ring (Fe–Nτ:

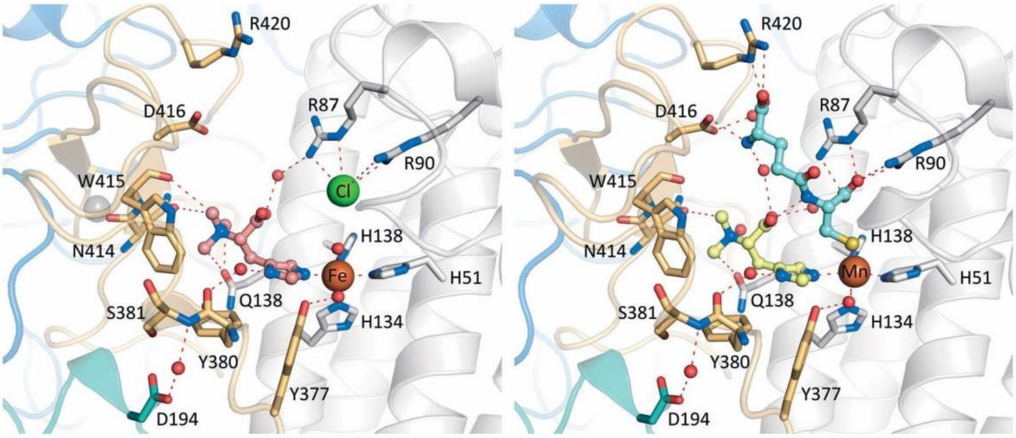


Figure 2. The active site of EgtB$_{\text{thermo}}$ in the ternary complex with TMH and iron (left), or with DMH, γGC, and manganese (right). Localized water molecules are shown in red. A chloride ion (green) in the ternary complex occupies the cationic docking site of γGC. Unbiased difference electron densities of bound ligands are shown in Figure S7.

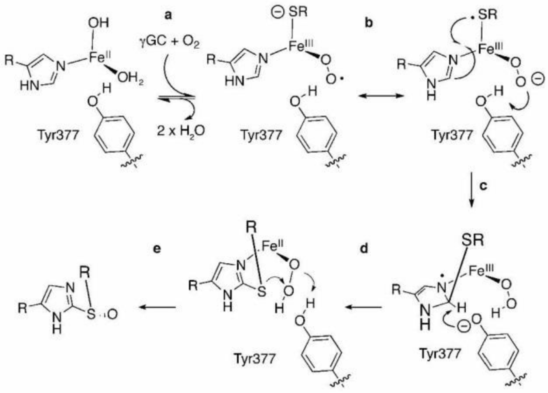
2.2 Å; Figure 2) and two water molecules (Fe–O: 2.1 and 2.2 Å) join the three-histidine facial triad in an octahedral coordination sphere around the iron center. The substrate imidazole ring also interacts with Tyr380 through a water-mediated hydrogen bond (5.4 Å), and the 1-carboxylate group is loosely connected to Arg87 (4.9 Å), again via a bridging water molecule. Two N-α-methyl groups of TMH pack against the indole side chain of Trp415. The third N-α-methyl group appears to make dipolar contacts to the amide side chains of Gln137 (3.2 Å) and Asn414 (3.5 Å).

To identify the binding mode of the second substrate γGC, we examined the structure of the quaternary complex between EgtB$_{\text{thermo}}$, DMH, γGC, and manganese(II). Initial attempts to soak the iron-containing binary complex with γGC resulted in disintegration of the crystals. As a solution to this problem, we used manganese-reconstituted EgtB$_{\text{thermo}}$. This enzyme complex is inactive[1d] but the corresponding crystals were tolerant to soaking with DMH and γGC and diffracted to a resolution of 1.98 Å (Figure 2 and Tables S2, S3). In this complex, the active-site residues and DMH adopt superimposable positions with respect to the ternary complex (r.m.s.d. = 0.041). γGC coordinates as the fifth ligand to the metal center (Mn–S: 2.6 Å). This direct metal–thiolate contact is consistent with the observed sulfur-to-iron charge-transfer transition absorption band in the γGC complex with ferric EgtB$_{\text{thermo}}$ (Figure S5). The α-amino group and the two carboxylates of γGC form salt bridges to Asp416 (2.8 Å), Arg420 (2.7 Å), Arg87 (3.0 Å), and Arg90 (2.6 Å). The amide function of γGC hydrogen bonds with the 1-carboxylate of DMH (2.8 Å).

To test whether the γGC-binding mode in the manganese complex properly reflects substrate binding by the iron-containing and therefore active form of EgtB$_{\text{thermo}}$, we probed γGC binding by site-directed mutagenesis. Specifically, we produced an EgtB$_{\text{thermo}}$ variant in which Asp416 is exchanged for Asn (EgtB$_{\text{D416N}}$). This mutation increases $K_{\text{M},\gamma\text{GC}}$ by 200-fold but does not significantly change $K_{\text{M,TMH}}$ or k_{cat}. In a complementary experiment, we assayed wild-type EgtB$_{\text{thermo}}$ and EgtB$_{\text{D416N}}$ with a γGC derivative that lacks the α-amino function (N-glutaryl cysteine, NGC). NGC is a 100-fold less efficient sulfur donor for wild type EgtB$_{\text{thermo}}$ but a 10-fold better substrate for EgtB$_{\text{D416N}}$ than γGC. Evidently, the salt bridge between the α-amino group of γGC and Asp416 found in the manganese-containing quaternary complex of EgtB$_{\text{thermo}}$ is also important for substrate recognition during catalysis.

The sixth metal ligand in the quaternary complex is a water molecule (Mn–O: 2.5 Å), which also hydrogen bonds to the phenolic side chain of Tyr377 (2.8 Å). Additionally, this coordination site maps to the mouth of a narrow tunnel connecting the active site with the protein exterior (Figure S8). It is tempting to view this tunnel as the path through which O_2 accesses the active site.

Based on the observation that γGC, TMH, and possibly O_2 are direct ligands to the catalytic iron center, we propose the following mechanism for EgtB-catalyzed sulfoxide synthesis (Scheme 2): Binding of γGC and O_2 to the EgtB



Scheme 2. Proposed mechanism for EgtB-catalyzed C–S bond formation and sulfoxidation.

complex with TMH results in an iron(III) superoxo species (step a). This species may have partial iron(III) character and forms a complex with a peroxide anion $(\mathrm{O_2}^{-2})$ and a thiyl radical (step b). The peroxide anion is stabilized by proton transfer from Tyr377 and the thiyl radical attacks the imidazole ring to form an iminyl radical (step c), which rearomatizes through deprotonation and ligand-to-metal electron transfer (step d). Sulfoxidation of the new thioether reduces the iron(II) peroxo species to return EgtB to its resting state (step e).

EgtB represents a novel type of non-heme iron enzyme, which catalyzes oxidative C–S bond formation and sulfoxidation. The crystal structure of EgtB provides the basis for testing mechanistic proposals by kinetic, spectroscopic, and computational methods[13] and also opens the door for protein engineering to generate tailor-made sulfur transferases. Our demonstration that mutation of a single residue (D416) suffices to change the sulfur donor specificity of $\mathrm{EgtB}_{\mathrm{thermo}}$ from γGC to NGC by almost three orders of magnitude bodes well for this objective. We anticipate that the structure of EgtB will also stimulate mechanistic investigations into the ovothiol biosynthetic sulfoxide synthase OvoA.[4]

Received: October 13, 2014

Keywords: enzyme catalysis · ergothioneine · non-heme iron enzymes · ovothiol · sulfur transfer

---

[1] a) D. B. Melville, W. H. Horner, C. C. Otken, M. L. Ludwig, *J. Biol. Chem.* **1955**, *213*, 61–68; b) D. B. Melville, D. S. Genghof, E. Inamine, V. Kovalenko, *J. Biol. Chem.* **1956**, *223*, 9–17; c) I. K. Cheah, B. Halliwell, *Biochim. Biophys. Acta* **2012**, *1822*, 784–793; d) F. P. Seebeck, *J. Am. Chem. Soc.* **2010**, *132*, 6632–6633.

[2] a) B. D. Paul, S. H. Snyder, *Cell Death Differ.* **2009**, *47*, S149–S149; b) T. Pluskal, M. Ueno, M. Yanagida, *PLoS One* **2014**, *9*, e97774; c) I. K. Cheah, R. L. Ong, J. Gruber, T. S. Yew, L. F. Ng, C. B. Chen, B. Halliwell, *Free Radical Res.* **2014**, *47*, 1036–1045; d) M. H. Bello, V. Barrera-Perez, D. Morin, L. Epstein, *Fungal Genet. Biol.* **2012**, *49*, 160–172; e) S. C. Emani, M. J. Williams, I. J. Wiid, N. F. Hiten, A. J. Viljoen, R. D. Pietersen, P. D. van Helden, B. Baker, *Antimicrob. Agents Chemother.* **2013**, *57*, 3202–3207.

[3] A. Vit, L. E. Misson, W. Blankenfeldt, F. P. Seebeck, *Acta Crystallogr. Sect. F* **2014**, *70*, 676–680.

[4] a) A. Braunshausen, F. P. Seebeck, *J. Am. Chem. Soc.* **2011**, *133*, 1757–1759; b) H. Song, A. S. Her, F. Raso, Z. Zheng, Y. Huo, P. Liu, *Org. Lett.* **2014**, *16*, 2122–2125; c) H. Song, M. Leninger, N. Lee, P. Liu, *Org. Lett.* **2013**, *15*, 4854–4857; d) G. T. Mashabela, F. P. Seebeck, *Chem. Commun.* **2013**, *49*, 7714–7716.

[5] a) C. A. Joseph, M. J. Maroney, *Chem. Commun.* **2007**, 3338–3349; b) P. L. Roach, I. J. Clifton, V. Fulop, K. Harlos, G. J. Barton, J. Hajdu, I. Andersson, C. J. Schofield, J. E. Baldwin, *Nature* **1995**, *375*, 700–704.

[6] a) B. Li, J. P. Yu, J. S. Brunzelle, G. N. Moll, W. A. van der Donk, S. K. Nair, *Science* **2006**, *311*, 1464–1467; b) E. Sasaki, X. Zhang, H. G. Sun, M. Y. Lu, M. Y. Liu, T. L. Liu, A. Ou, J. Y. Li, Y. H. Chen, S. E. Ealick, H. W. Liu, *Nature* **2014**, *510*, 427–431; c) A. Chatterjee, N. D. Abeydeera, S. Bale, P.-J. Pai, P. C. Dorrestein, D. H. Russell, S. E. Ealick, T. P. Begley, *Nature* **2011**, *478*, 542–U146; d) F. Berkovitch, Y. Nicolet, J. T. Wan, J. T. Jarrett, C. L. Drennan, *Science* **2004**, *303*, 76–79; e) P. L. Roach, I. J. Clifton, C. M. Hensgens, N. Shibata, C. J. Schofield, J. Hajdu, J. E. Baldwin, *Nature* **1997**, *387*, 827–830; f) D. H. Scharft, P. Chankhamjon, K. Scherlach, T. Heinekamp, K. Willing, A. A. Brakhage, C. Hertweck, *Angew. Chem. Int. Ed.* **2013**, *52*, 11092–11095; *Angew. Chem.* **2013**, *125*, 11298–11301.

[7] J. D. Gardner, B. S. Pierce, B. G. Fox, T. C. Brunold, *Biochemistry* **2010**, *49*, 6033–6041.

[8] a) S. A. McMahon, J. L. Miller, J. A. Lawton, D. E. Kerkow, A. Hodes, M. A. Marti-Renom, S. Doulatov, E. Narayanan, A. Sali, J. F. Miller, P. Ghosh, *Nat. Struct. Mol. Biol.* **2005**, *12*, 886–892; b) T. Dierks, A. Dickmanns, A. Preusser-Kunze, B. Schmidt, M. Mariappan, K. von Figura, R. Ficner, M. G. Rudolph, *Cell* **2005**, *121*, 541–552.

[9] J. Le Coq, P. Ghosh, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14649–14653.

[10] a) K. Nagata, J. Ohtsuka, M. Takahashi, A. Asano, H. Iino, A. Ebihara, M. Tanokura, *Proteins Struct. Funct. Bioinf.* **2008**, *70*, 1103–1107; b) S. S. Rajan, X. Yang, L. Shuzvalova, F. Collart, W. F. Anderson, *Biochemistry* **2004**, *43*, 15472–15479; c) D. R. Cooper, K. Grelewska, C.-Y. Kim, A. Joachimiak, Z. S. Derewenda, *Acta Crystallogr. Sect. F* **2010**, *66*, 219–224.

[11] a) G. L. Newton, S. S. Leuung, J. L. Wakabayashi, M. Rawat, R. C. Fahey, *Biochemistry* **2011**, *50*, 10751–10760; b) J. Feng, Y. Che, J. Milse, L. Yin, C. Rückert, X. H. Shen, S. W. Qi, J. Kalinowski, S. J. Liu, *J. Biol. Chem.* **2006**, *281*, 10778–10785; c) T. T. Liu, N. Y. Zhou, *J. Bacteriol.* **2012**, *194*, 3987–3994.

[12] J. Penner-Hahn, *Curr. Opin. Chem. Biol.* **2007**, *11*, 166–171.

[13] E. A. Bushnell, G. B. Fortowsky, J. W. Gauld, *Inorg. Chem.* **2012**, *51*, 13351–13356.

12.3 Structure of the Ergothioneine Biosynthetic Amidohydrolase EgtC

A five-gene cluster *egtABCDE*-encoded EgtC is specific for Actinobacteria, and for few cyanobacterial and proteobacterial species (Jones *et al.*, 2014). Most other species, especially fungi, lack an *egtA*, *egtC* and/or *egtE*. The function of these genes might be complemented by other genes located elsewhere in the genome. Hence, the EgtC-catalyzed reaction can mainly be considered as an actinobacterial trait.

EgtC from *M. smegmatis* is a 25 kDa enzyme that catalyzes the hydrolysis of γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide (QHC) to glutamate and N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide. For this purpose, EgtC utilizes its N-terminal cysteine residue C2 as a nucleophile and is therefore classified as an Ntn amidohydrolase (Pfam CL0052) (Brannigan *et al.*, 1995). Despite the similarities in reactivity and a common fold, members of the Ntn hydrolase superfamily do not necessarily share recognizable sequence homology (Oinonen & Rouvinen, 2000). However, for most members of this family, a conformational change upon substrate binding to the enzyme is suggested (Isupov *et al.*, 1996). After having determined the crystal structure of apo EgtC from *M. smegmatis* via selenium SAD (Vit, 2010), cocrystallization of EgtC with glutamine was persued. Additionally, cocrystallization of an inactive EgtC variant EgtC_{C2A} with γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide, the product of the EgtB-catalyzed reaction and substrate to EgtC, was attempted. Such a complex would shed light on substrate binding, the reaction mechanism and the stereochemistry of the sulfoxide moiety of γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide. For this purpose, the isopeptide substrate had to be isolated from an *in vitro* reaction mixture of recombinant enzymes and their substrate via HPLC (performed by G. T. Mashabela). Subsequently, cocrystallization of EgtC with glutamine and EgtC_{C2A} with γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide yielded a high resolution crystal structure which has been determined by MR using the apo structure of EgtC (1.75 Å). Data collection and refinement statistic of the EgtC structures are given in Table 41 and Table 42. These EgtC structures reveal the overall topology of this Ntn hydrolase, as well as the key interactions between the isopeptide substrate QHC and the enzyme. Most importantly, the stereochemistry at the sulfoxide position of γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide can be determined using the high resolution complex of EgtC_{C2A} and QHC.

Table 41: Data collection statistics for *M. smegmatis* EgtC apo structure and complexes. Values in parentheses are for the highest resolution shell. Gln = glutamine. QHC = γ-glutamyl-N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide. EgtC$_{C2A}$ = inactive variant.

Dataset	Se SAD[+] EgtC_apo	EgtC$_{wt}$_apo	EgtC$_{wt}$_Gln	EgtC$_{C2A}$_QHC
Wavelength (Å) / beamline[≠]	0.9786 / SLS, PXII	0.9792 / SLS, PXII	0.9184 / BESSY II, 14.1	1.000 / PETRA III, P11
Resolution range (Å)	47.83-1.85 (1.88-1.85)	47.64-1.75 (1.78-1.75)	46.44-1.82 (1.85-1.82)	49.31-1.7 (1.73-1.70)
Space group	$P2_1$	$P2_1$	$P2_12_12_1$	$P2_1$
Unit cell parameters (Å) (°)	129 69.4 160.4 90 94.73 90	129 69.5 160.4 90 94.84 90	61.8 110 139.3 90 90 90	128.5 69.4 159.4 90 94.67 90
Mosaicity (°)[†]	0.150	0.170	0.105	0.181
Total No. of measured reflections	3248829 (162534)	958644 (47488)	355388 (18637)	2062156 (99295)
Unique reflections	236707 (11572)	283662 (14038)	85790 (4472)	299928 (14544)
Multiplicity	6.9 (7.0)	3.4 (3.4)	4.1 (4.2)	6.9 (6.8)
Mean I/σ(I)	15.9 (3.1)[§]	10.3 (1.9)	11.5 (2.0)	16.9 (2.1)
Completeness (%)	98.3 (97.2)	99.6 (99.8)	99.9 (99.9)	97.8 (96.2)
R_{meas} (%)[‡]	14.2 (112.4)	8.8 (77.7)	14.1 (91.4)	7.7 (104.6)
$R_{p.i.m.}$ (%)[$]	5.3 (41.8)	4.7 (41.5)	6.9 (44.5)	2.9 (39.7)

[+] Data collection statistics for unmerged Friedel pairs.

[†] Mosaicity values reported by *XDS* (Kabsch, 2010).

[§] The anomalous signal is significant to 4.6 Å resolution according to *Xtriage* from *PHENIX* (Zwart *et al.*, 2005; Evans, 2011). As the resolution was significantly higher, the data was cut at mean I/σ(I) = 3.1 in the highest resolution shell.

[≠] SLS: Swiss Light Source (Paul Scherrer Institute, Villigen, CH), BESSY: Berlin Electron Storage Ring Society for Synchrotron Radiation (Helmholtz Zentrum Berlin, Berlin, D), PETRAIII: Positron-Elektron-Tandem-Ring-Anlage III at DESY (Deutsche Elektronen-Synchrotron, Hamburg, D).

[‡] $R_{meas} = \Sigma_{hkl} (N/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where N is the number of observations of the reflection with index hkl and I_i is the intensity of its ith observation.

[$] $R_{p.i.m.} = \Sigma_{hkl} (1/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$ (Weiss, 2001).

Table 42: Refinement statistics for *M. smegmatis* EgtC apo structure and complexes. Values in parentheses are for the highest resolution shell. Gln = glutamine. QHC = γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide. $EgtC_{C2A}$ = inactive variant.

Dataset	$EgtC_{wt}$_apo	$EgtC_{wt}$_Gln	$EgtC_{C2A}$_QHC
Resolution range (Å)	47.64-1.75 (1.78-1.75)	35.46-1.82 (1.84-1.82)	49.31-1.7 (1.73-1.70)
R_{work} (%)	14.9 (22.9)	14.7 (24.4)	14.4 (24.2)
R_{free} (%)	18.0 (25.8)	17.8 (28.2)	17.1 (27.5)
No. of non-H atoms			
Protein	20469	6810	20579
Ion	-	-	-
Ligand	419	40	396
Water	2481	1224	3083
R.m.s. deviations			
Bonds (Å)	0.006	0.006	0.009
Angles (°)	1.176	1.037	1.246
Average B factors (Å^2)			
Protein	24	15	24
Ion	-	-	-
Ligand	48	10	38
Water	34	32	34
Ramachandran plot			
Favored regions (%)	96.9	97.7	97.2
Outliers (%)	0.2	0.1	0
MolProbity score[#]	1.45	0.97	1.26
PDB entry code	-	-	-

[#] As reported by *MolProbity* (Chen *et al.*, 2010).

EgtC forms tetramers in the asymmetric units of the crystals (Figure 78A). The tetramer probably resembles the physiological quarternary structure of EgtC according to the buried surface area and oligomer predictions by the PISA server (Krissinel & Henrick, 2007). Similarly, EgtC elutes as a tetramer from analytical size exclusion chromatography. The substrate of EgtC coelutes with the protein, demonstrating that it specifically binds to EgtC (Figure 78B). Interestingly, at least dimerization of EgtC appears obligatory for activity because the substrate binding site is formed by residues from both molecules of the dimer (Figure 78A).

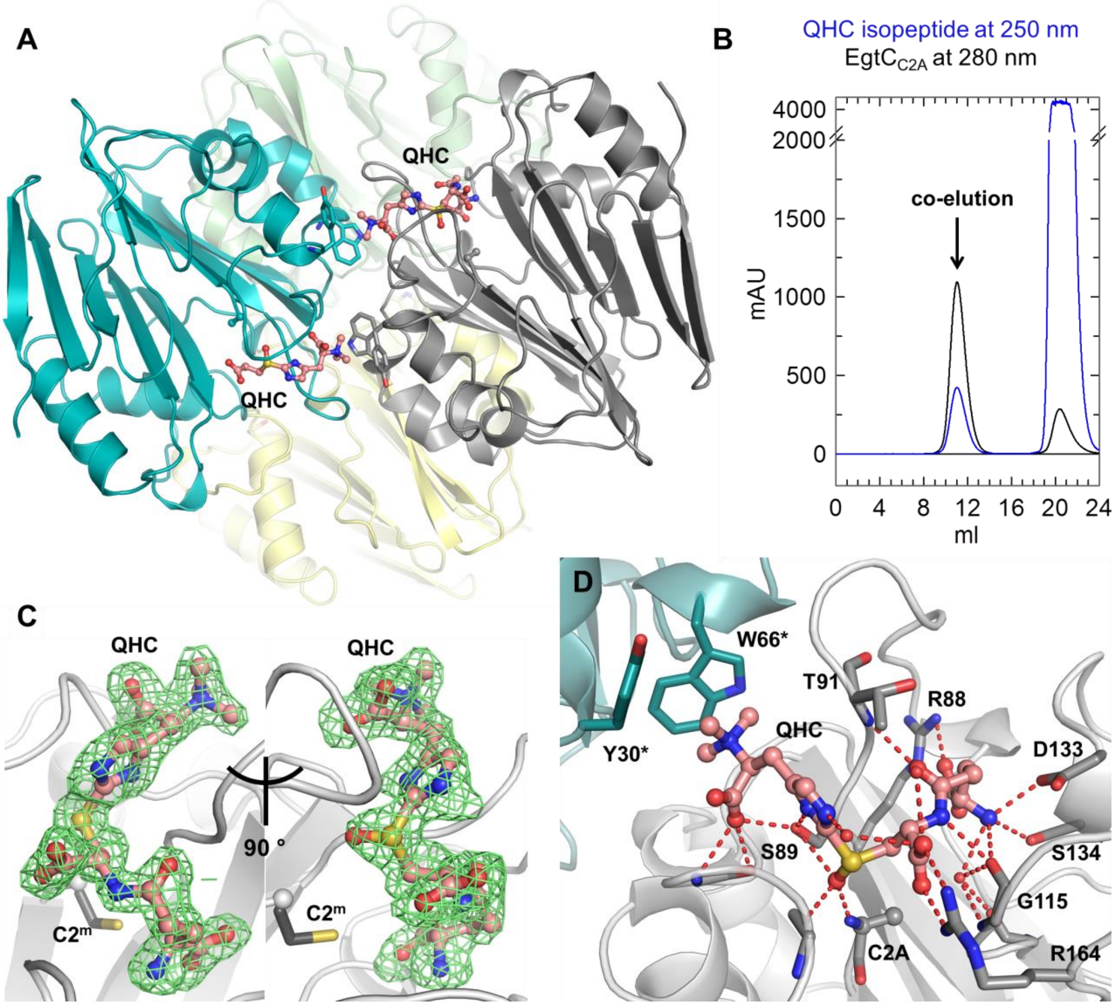


Figure 78: The crystal structure of EgtC$_{C2A}$ in complex with γ-glutamyl-N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide (QHC). A: EgtC is at least an obligate dimer (yellow/teal), because QHC (salmon) binding involves two residues from the second EgtC molecule. However, a tetramer is more likely because of the amount of buried surface. B: QHC coeluates with EgtC$_{C2A}$ from an S200 10/300 GL size exclusion chromatography column. C: The presence of QHC in the EgtC$_{C2A}$ active site is clearly indicated by positive m|F$_{obs}$|-D|F$_{calc}$| electron density (σ-level = 2). This density could be used to determine the stereochemistry of the sulfoxide moiety of QHC. C2^{m} is modeled from the wild type protein. D: QHC is recognized by a dense array of hydrogen bonds.

The tertiary structure of EgtC is formed by a four layer α/β assembly with two central antiparallel 5- and 6-stranded β-sheets. The characteristic nucleophilic residue of the Ntn hydrolase EgtC is provided by the 1,2-aminothiol function of the N-terminal C2. Removal of the preceeding residue M1 is probably a result of bacterial methionine aminopeptidase activity (Lowther & Matthews, 2000). In the 1.7 Å resolution structure of EgtC$_{C2A}$ with γ-glutamyl-N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide, positive m|F$_{obs}$|-D|F$_{calc}$| electron density clearly confirms the presence of the substrate in the binding pocket (Figure 78C). This electron density can be used to elucidate the stereochemical center of the sulfoxide moiety of γ-glutamyl-N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide. This sulfoxide stereochemistry in turn sheds light on the catalytic mechanism of the sulfoxide synthase EgtB, also investigated in this study. In the complex of EgtB with γ-glutamy-cysteine and N^{α},N^{α},N^{α}-trimethyl-histidine, the predicted O_2 binding site makes a clear prediction as to which sulfoxide stereoisomer should be generated if suloxidation occurs according to this mechanism (Goncharenko *et al.*, 2015). The observed stereochemistry of the substrate in the $EgtC_{C2A}$ complex does indeed show this predicted stereochemistry and demonstrates that the sulfoxide center is in *S* configuration (Figure 78C). γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide is bound by a dense array of hydrogen bond interactions in the active center of EgtC (Figure 78D). The amino acid moiety of the glutamyl part of γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide is stabilized by R88 (carboxyl group), D133 and S134 (amino group) and the backbone of G115. The carboxyl group of the cysteine moiety is stabilized by R164, and the sulfoxide moiety establishes a hydrogen bond to the N-terminal amino group of EgtC. The trimethylated amino group of γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide is recognized by Y30* and W66* of the other EgtC molecule of the obligate EgtC homodimer.

In all crystal structures, EgtC and $EgtC_{C2A}$ adopt very similar conformations and overall structures. The only flexible segment is a loop from A90 to P95 above the active site. This loop can adopt an open ore a closed conformation and generates a more open or closed active site. In the crystal structure of EgtC with glutamine, the flexible loop adopts a closed conformation, while the crystal structure of $EgtC_{C2A}$ and γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide displays an open loop conformation (Figure 79A). The latter conformation and mode of substrate binding seem to be unproductive, because the cysteine nucleophile is too far away from the carbonyl carbon at which the nucleophile should attack (Figure 79A). However, if both structures of EgtC with glutamine and $EgtC_{C2A}$ with γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide are superposed, an active conformation of the protein can be derived and modeled, in which the cysteine nucleophile is in close proximity of the carbonyl carbon (Figure 79B).

The proposed active conformation of EgtC indicates that the flexible A90 to P95 loop is probably able to push γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide into the active site of EgtC, mainly by T91, thereby allowing cytalysis of the hydrolysis reaction. This binding mode and loop conformation would be productive, and fit to the hypothesis that enzymes of this family undergo a conformational change upon substrate binding (Isupov *et al.*, 1996).

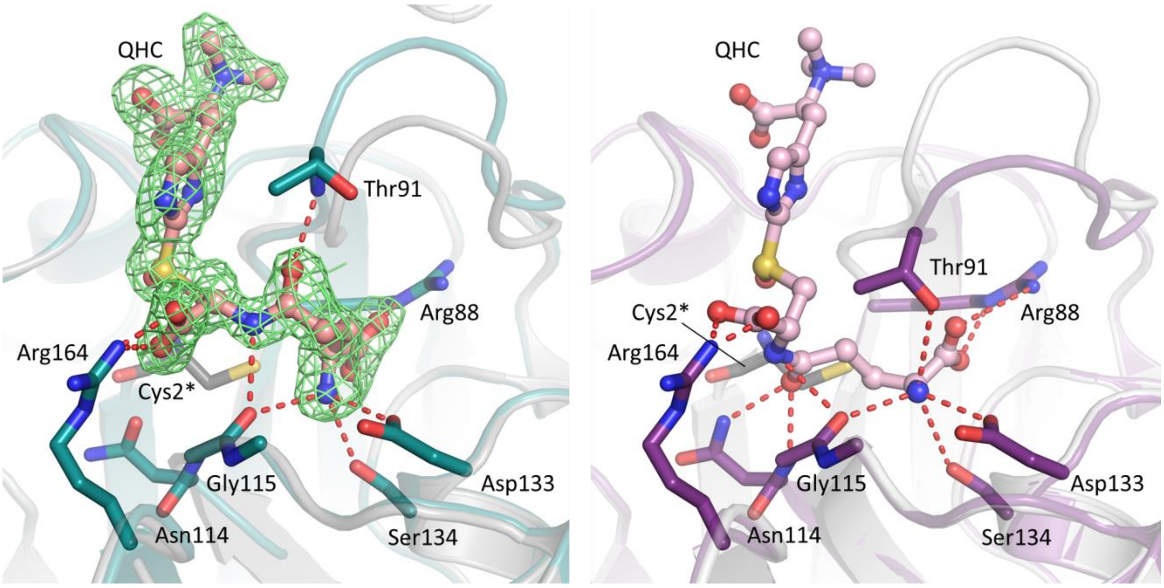


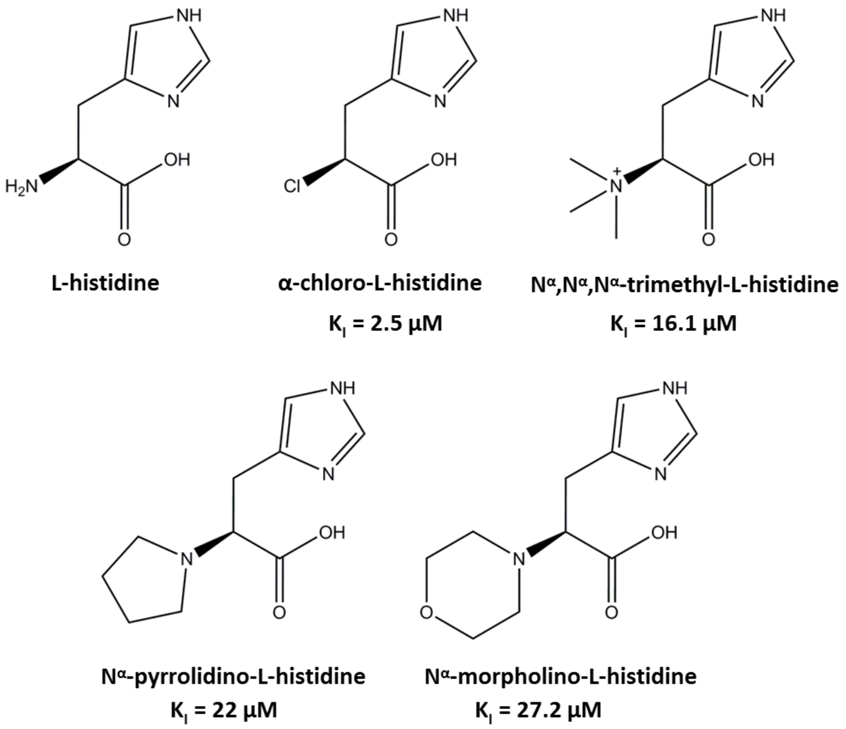
Figure 79: Substrate binding to EgtC. A: Superposition of the EgtC glutamine complex (grey) with the EgtC$_{C2A}$ γ-glutamyl-N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide (QHC) complex. The flexible A90-P95 loop is located on top of the active site and adopts an open (teal) and a closed (grey) conformation. In this QHC complex, the nucleophilic C2 would be too far away from the carbonyl carbon of QHC to hydrolyse the substrate. B: Model of a productive binding mode of QHC to EgtC (purple). The glutamyl moiety of QHC adopts the conformation of glutamine form the EgtC glutamine complex, leading to a more productiove binding mode.

EgtC and EgtA, the enzyme producing γ-glutamyl-cysteine, have probably been recruited into actinobacterial ergothioneine biosynthesis is that these enzymes allow the bacteria to use γ-glutamyl-cysteine rather than cysteine for ergothioneine production. The use of γ-glutamyl-cysteine is advantageous because it avoides high intracellular cysteine concentrations, which might cause production of oxygen radicals catalyzed by cysteine-chelated transition metals (Fahey, 2013). Hence, γ-glutamyl-cysteine is a safer substrate in ergothioneine biosynthesis

Taken together, EgtC is the first enzyme of the ubiquitous GATase_6 protein family (PF13522) with a defined catalytic activity and physiological function described. The stereochemistry of the sulfoxide moiety of γ-glutamyl-N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide is in agreement with data on EgtB previously published (Goncharenko *et al.*, 2015).

12.4 Inhibiton of Ergothioneine Biosynthetic Methyltransferase EgtD Using Histidine Derivatives

Since ergothioneine biosynthesis might be a drug target for fighting *M. tuberculosis* and the respective disease, potential inhibitors for EgtD were synthesized (S. Kostic and R. Burns). These compounds (Figure 80) were analyzed in a coupled enzymatic EgtD assay (Dorgan *et al.*, 2006; Vit *et al.*, 2015) and the inhibition constant K_I was determined for every compound.



L-histidine

α-chloro-L-histidine
K_I = 2.5 µM

N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-L-histidine
K_I = 16.1 µM

N$^{\alpha}$-pyrrolidino-L-histidine
K_I = 22 µM

N$^{\alpha}$-morpholino-L-histidine
K_I = 27.2 µM

Figure 80: Histidine and potential histidine-derived ergothioneine biosynthesis inhibitors. The potential inhibitors were synthesized (by S. Kostic and R. Burns) and tested in a coupled assay (by L. Misson). The K_I of N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidine, the product of the EgtD-catalyzed reaction, was determined as 16.1 µM. Hence, the final product N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidine affects EgtD by product inhibition. N$^{\alpha}$-pyrrolidino-histidine and N$^{\alpha}$-morpholino-histidine are less effective with K_Is of 22 µM and 27.2 µM, respectively. The most potent EgtD inhibitor is α-chloro-histidine with a K_I of 2.5 µM.

In order to investigate the exact positions of the amino group substituents when bound to EgtD, apo crystals of EgtD were produced (Vit *et al.*, 2014) and soaked with 20 mM of inhibitor in reservoir solution for 30 min. Then, diffraction data were collected on beamline 14.1 at the BESSY II synchrotron (HZB, Berlin, Germany). The inhibitor complexes were determined via rigid body refinement with the EgtD apo structure. The data collection and refinement statistics are depicted in Table 43 and Table 44.

Table 43: Data collection statistics for *M. smegmatis* EgtD inhibitor complexes. Values in parentheses are for the highest resolution shell. AVJ = N^{α},N^{α},N^{α}-trimethyl-histidine, AVK = α-chloro-histidine, AVM = N^{α}-morpholino-histidine, AVO = N^{α}-pyrrolidino-histidine.

Dataset	EgtD_AVJ	EgtD_AVK	EgtD_AVM	EgtD_AVO
Wavelength (Å) / beamline≠	0.9184/ BESSY II, 14.1	0.9184 / BESSY II, 14.1	0.9184 / BESSY II, 14.1	0.9184 / BESSY II, 14.1
Resolution range (Å)	49.33-1.75 (1.78-1.75)	49.34-1.83 (1.87-1.83)	48.85-1.85 (1.89-1.85)	49.25-1.9 (1.94-1.9)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	71.0, 79.4, 136.0	71.2, 79.2, 136.3	71.7, 76.2, 138.7	71.8, 79.5, 139.0
(°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Mosaicity (°)†	0.071	0.107	0.131	0.089
Total No. of measured reflections	1046638 (58915)	897454 (43155)	862107 (46234)	822801 (54415)
Unique reflections	78334 (4258)	68689 (4158)	65608 (4003)	61947 (3955)
Multiplicity	13.4 (13.8)	13.1 (10.4)	13.1 (11.5)	13.3 (13.8)
Mean I/σ(I)	15.3 (1.9)	17.3 (2.1)	18.0 (1.9)	19.7 (2.3)
Completeness (%)	100.0 (100.0)	99.9 (99.1)	100.0 (100.0)	100.0 (100.0)
R_{meas} (%)‡	14.5 (168.5)	11.4 (114.5)	11.7 (146.9)	8.9 (143.4)
$R_{p.i.m.}$ (%)$	3.9 (44.8)	3.1 (34.8)	3.2 (42.7)	2.4 (38.2)

†Mosaicity values reported by *XDS* (Kabsch, 2010).

≠ BESSY: Berlin Electron Storage Ring Society for Synchrotron Radiation (Helmholtz Zentrum Berlin, Berlin, D).

‡$R_{meas} = \Sigma_{hkl} (N/(N-1))^{1/2} \Sigma_i \mid I_i (hkl) - < I (hkl) > \mid / \Sigma_{hkl} \Sigma_i I_i (hkl)$, where N is the number of observations of the reflection with index hkl and I_i is the intensity of its i^{th} observation.

$$R_{p.i.m.}$. $= \Sigma_{hkl} (1/(N-1))^{1/2} \Sigma_i \mid I_i (hkl) - < I (hkl) > \mid / \Sigma_{hkl} \Sigma_i I_i (hkl)$ (Weiss, 2001).

Table 44: Refinement statistics for *M. smegmatis* EgtD inhibitor complexes. Values in parentheses are for the highest resolution shell. AVJ = N$^\alpha$,N$^\alpha$,N$^\alpha$-trimethyl-histidine, AVK = α-chloro-histidine, AVM = N$^\alpha$-morpholino-histidine, AVO = N$^\alpha$-pyrrolidino-histidine.

Dataset	EgtD_AVJ	EgtD_AVK	EgtD_AVM	EgtD_AVO
Resolution range (Å)	49.33-1.75 (1.78-1.75)	49.34-1.83 (1.87-1.83)	48.85-1.85 (1.89-1.85)	49.25-1.9 (1.94-1.9)
R_{work} (%)	15.1 (23.1)	15.7 (25.9)	16.3 (25.3)	17.0 (28.9)
R_{free} (%)	17.8 (27.2)	18.3 (29.9)	18.4 (25.4)	20.7 (37)
No. of non-H atoms				
Protein	5046	4955	4929	4940
Ion	1	1	-	-
Ligand	70	48	44	30
Water	732	525	479	309
R.m.s. deviations				
Bonds (Å)	0.013	0.006	0.004	0.004
Angles (°)	1.6	1.01	0.93	0.876
Average B factors (Å^2)				
Protein	22	28	28	41
Ion	17	20	-	-
Ligand	30	34	34	35
Water	37	36	37	45
Ramachandran plot				
Favored regions (%)	98.6	98.5	98.3	98.3
Outliers (%)	0	0.2	0	0
MolProbity score[#]	1.07	0.87	0.75	0.78
PDB entry code	-	-	-	-

[#] As reported by *MolProbity* (Chen *et al.*, 2010)

The crystal structures of EgtD inhibitor complexes were obtained by soaking of the compounds into native apo crystals obtained in the original apo condition reported (Vit *et al.*, 2014). All complexes were determined at high resolution (1.75-1.9 Å) and refined to reasonable R-factors (Table 44). The $m|F_{obs}|-D|F_{calc}|$ electron densities for α-chloro-histidine and N$^\alpha$,N$^\alpha$,N$^\alpha$-trimethyl-histidine were well-defined and interpretable after initial rigid body refinement. However, $m|F_{obs}|-D|F_{calc}|$ densities for the N$^\alpha$-pyrrolidino and N$^\alpha$ morpholino derivatives were not clearly visible initially. This was caused by a slightly more open conformation of EgtD induced by the N$^\alpha$-pyrrolidino and N$^\alpha$ morpholino-moieties of the histidine derivatives. The conformational change is also seen in the r.m.s.d values of 0.369 and 0.401 of the N^{α}-pyrrolidino-histidine and N^{α}-morpholino-histidine complexes, respectively. These r.m.s.d. values are higher than the r.m.s.d. of the α-chloro-histidine and the N^{α},N^{α},N^{α}-trimethylhistidine complex (0.136). For this reason, the residues in the N^{α}-pyrrolidino-histidine and N^{α}-morpholino-histidine complexes had to be slightly rearranged into the density. Only then, the $m|F_{obs}|-D|F_{calc}|$ density for N^{α}-pyrrolidino- and N^{α}-morpholino-histidine was clearly visible and well defined. All inhibitors were refined to reasonable average B-factors of 30 Å^2 to 35 Å^2, and detailed views, including $m|F_{obs}|-D|F_{calc}|$ electron densities, are shown in Figure 81.

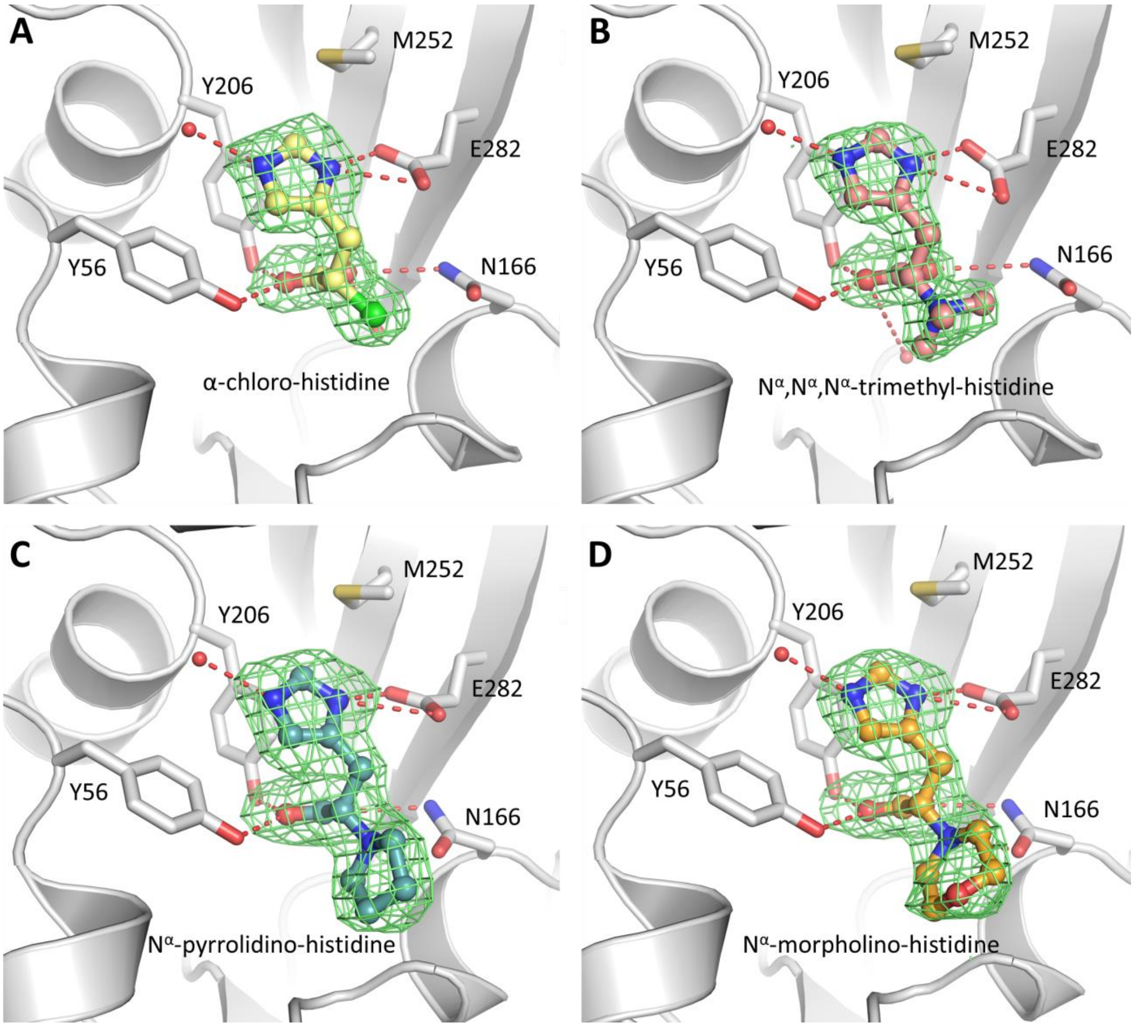


Figure 81: Crystal structures of EgtD with potential inhibitors. Unbiased $m|F_{obs}|-D|F_{calc}|$ electron densities (σ-level = 2) of the compounds confirm their presence in the crystal structures. Typical interactions of the imidazole moiety of the histidine derivatives with M252 and E282 of EgtD are present in all four structures. A: EgtD with α-chloro-histidine. B: EgtD with N^{α},N^{α},N^{α}-trimethyl-histidine. C: EgtD with N^{α}-pyrrolidino-histidine. The pyrrolidino group is in boat conformation. D: EgtD with N^{α}-morpholino-histidine. The morpholino group is in boat conformation.

The crystal structures of EgtD in complex with four different inhibitors demonstrate an identical binding mode for the imidazole moiety of histidine as the original EgtD structure. This binding mode, involving a water molecule, M252 and E282, is not surprising because no changes were made to this imidazole moiety, and EgtD is optimized to bind this moiety with high affinity. The same is true for the carboxyl group which is establishes hydrogen bonds to the hydroxyl groups of Y56 and Y206, as well as to the amide side chain of N166. However, modification of histidine at its α-amino group is very flexible, and the enzyme is also able to harbor bulky modifications such as the N^{α}-pyrrolidino or N^{α}-morpholino moieties, which are both present in boat conformation in EgtD. Interestingly, these modifications point into the SAM binding site of EgtD. Hence, EgtD, not binding SAM or SAH, could harbor even larger α-amino group modifications than the N^{α}-morpholino moiety. However, the N^{α}-morpholino- and N^{α}-pyrrolidino-histidine derivatives display less inhibitory potential in the EgtD assay than the derivatives with smaller α-amino group substituents. The K_I's of N^{α}-morpholino-histidine and N^{α}-pyrrolidino-histidine were determined as 22.0 µM and 27.2 µM, respectively, while the K_I's of α-chloro-L-histidine and the product of EgtD, N^{α},N^{α},N^{α}-trimethyl-histidine, were determined as 2.5 µM and 16.1 µM, respectively (L. Misson, unpublished data). For this reason, α-chloro-histidine is the most potent EgtD inhibitor tested in the assay, and it is more effective than N^{α},N^{α},N^{α}-trimethyl-histidine. This is surprising, because it means that the positive charge of the α-amino group of histidine is not crucial for substrate recognition by EgtD.

13 Synopsis and Outlook

Ergothioneine is a sulfur derivative of histidine betaine and plays important physiological roles in its producing organisms, for instance Mycobacteria, but also in humans who are not able to produce ergothioneine. In order to investigate the mycobacterial biosynthesis of ergothioneine, the SAM-dependent methyltransferase EgtD, the sulfoxide synthase EgtB and the Ntn amidohydrolase EgtC were examined using X-ray crystallography and biophysical as well as biochemical methods in collaboration with the group of Prof. Dr. Florian P. Seebeck from the University of Basel in Switzerland. Taken together, these three ergothioneine biosynthesis enzymes were successfully characterized by structural biology and biophysical as well as biochemical approaches.

13.1 Methyltransferase EgtD

In the first step of ergothioneine biosynthesis, the SAM-dependent methyltransferase EgtD catalyzes the three consecutive methyl transfers from SAM to L-histidine to produce N^{α},N^{α},N^{α}-trimethyl-histidine. In order to understand the catalytic activity and substrate specificity of EgtD, its apo structure was determined earlier via selenium SAD (Vit, 2010; Vit *et al.*, 2014). In this study, complexes of EgtD with N^{α},N^{α}-dimethyl-histidine and EgtD with N^{α},N^{α}-dimethyl-histidine and SAH have been determined.

The structure of EgtD revealed a conserved Rossman-fold domain including a seven-stranded β-sheet (Figure 82A) and a second domain formed by an N-terminal extension and an insertion after the fifth β-strand of the central β-sheet. This domain is unrelated to any other known methyltransferase domain and is important for substrate specificity. The cleft between the two domains harbors the catalytic center in which histidine and SAM/SAH are bound (Vit *et al.*, 2015). Upon SAM binding, a slight conformational change closes the active site of EgtD. The same conformational change was observed in other crystal structures of EgtD which were published recently (Jeong *et al.*, 2014).

EgtD catalyzes the trimethylation of histidine with remarkable apparent processivity (Seebeck, 2010). This processivity is a result of increased affinity towards methylated substrates as found in ITC experiments and this increased affinity explains the processivity of the EgtD-catalyzed reaction (Vit, 2010; Vit *et al.*, 2015). The processivity of the

EgtD-catalyzed reaction is important to channel a distinct fraction of the cellular histidine pool into ergothioneine biosynthesis without interfering with protein biosynthesis (Vit *et al.*, 2015).

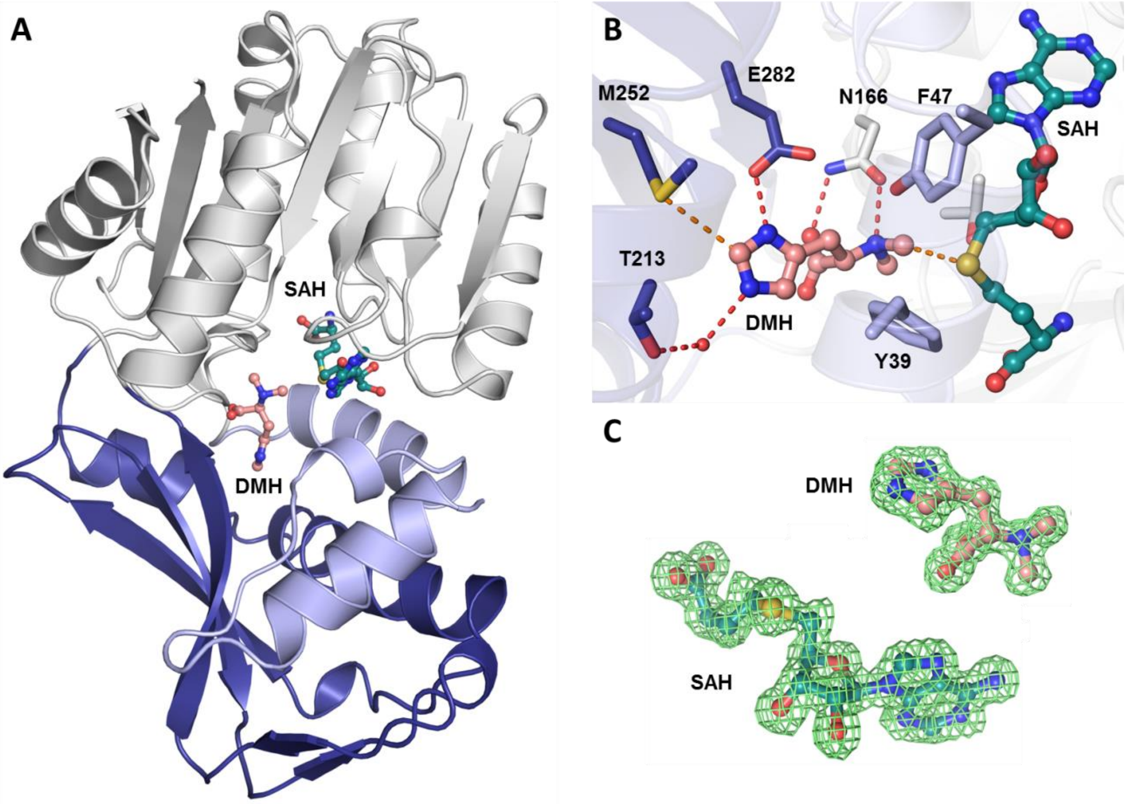


Figure 82: A: 1.51 Å crystal structure of EgtD with N$^{\alpha}$,N$^{\alpha}$-dimethyl-histidine (DMH) and S-adenosyl-homocysteine (SAH). B: Active site of EgtD. Residues M252 and E282 form the binding pocket for the imidazole moiety together with a water molecule coordinated by T213. N166 stabilizes the amino acid function of DMH. One methyl group is in contact with the sulfur atom of SAH. C: m|F_{obs}|-D|F_{calc}| electron densities (σ-level = 2.5) of SAH (left) and DMH (right).

The structural basis for substrate specificity and processivity of the EgtD-catalyzed reaction is a result of the dense network of interactions between the dimethylated α-amino group of histidine and EgtD (Y39, F47 and N166, Figure 82B) (Vit *et al.*, 2015). Additionally, one N$^{\alpha}$-methyl group is in van-der-Waals contact with the sulfur atom of SAH, and N$^{\alpha}$,N$^{\alpha}$-dimethyl-histidine establishes a hydrogen bond between Nτ of the imidazole moiety and E282 of EgtD.

A bioinformatic analysis of mycobacterial and fungal genomes revealed that fungi, in contrast to Mycobacteria, often possess an additional EgtD homolog. This additional fungal homolog shows large variability in binding site positions 213, 252 and 282 (EgtD numbering) (Vit *et al.*, 2015). On the basis of this result, a tyrosine betaine synthase (Ybs) methyltransferase was identified in *Aspergillus nidulans*. Additionally, a tryptophan betaine synthase (HyoA) was identified in the wood-rotting fungus *Dichomitus squalens*. Since purification of HyoA was unsuccessful, a tryptophan-specific variant EgtD$_{\mathrm{M252V,E282A}}$ was successfully engineered and yielded an enzyme with an active site configuration similar to that in the fungal protein.

The structure of EgtD$_{\mathrm{M252,E282}}$ with tryptophan and SAH displays the aromatic amino acid in the enlarged binding site (Figure 83B). The mode of binding is very similar to the one of N$^{\alpha}$,N$^{\alpha}$-dimethyl-histidine and EgtD, but distinct mutations allow the enzymes to be highly specific for one type of aromatic amino acid. In contrast to EgtD (T213, M252, E282), tryptophan betaine synthases possess residues T213, V252 and A282 and thus they own a larger binding pocket to harbor the indole side chain of tryptophan. Tyrosine betaine synthases such as Ybs from *A. nidulans* probably utilize N213, Q252 and V282 for tyrosine binding (Figure 83C).

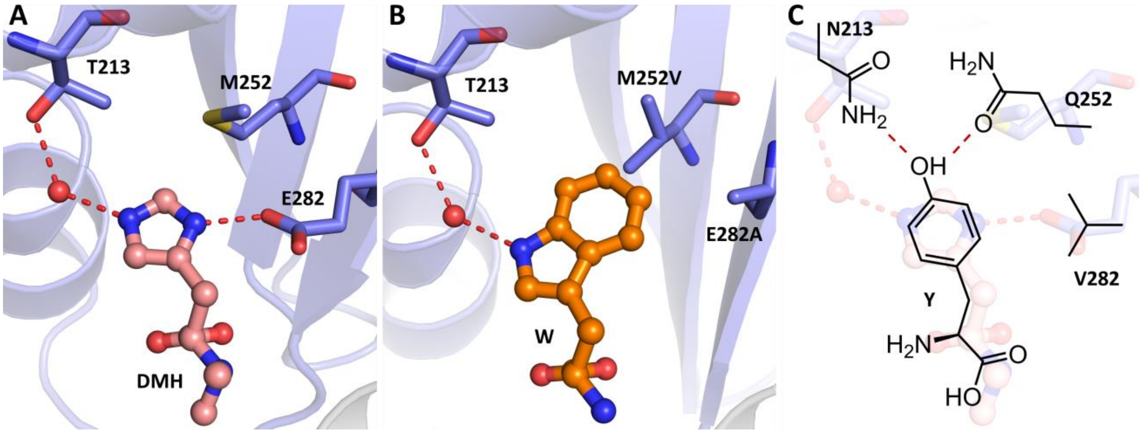


Figure 83: Active centers of aromatic amino acid betaine synthases (residue numbering is based on the EgtD sequence). A: EgtD coordinates the imidazole moiety of the histidine substrates by hydrogen bond interactions between Nτ and E282 and a water mediated bridge between Nπ and T213. B: EgtD$_{\mathrm{M252V,E282A}}$ possesses a larger binding pocket that is able to harbor the indole side chain of tryptophan via hydrophobic interactions and a water mediated hydrogen bond between the nitrogen atom of tryptophan and T213. C: Proposed binding interactions for tyrosine-specific betaine synthases, for instance Ybs from *A. nidulans*. The hydroxyl group of the tyrosine side chain probably establishes by two hydrogen bonds to N213 and Q252.

Taken together, the data obtained in this study do not only characterize the first step of ergothioneine biosynthesis, but also reveal the structural basis of aromatic amino acid betaine biosynthesis, which is distinct from glycine betaine biosynthesis. Additionally, aromatic amino acid betaines seem to be common metabolites also in fungi.

13.2 Sulfoxide Synthase EgtB

The second step in ergothioneine biosynthesis is the O_2-dependent carbon-sulfur bond formation and oxidation between the imidazole ring of N^{α},N^{α},N^{α}-trimethyl-L-histidine and the thiol side chain of γ-glutamyl-cysteine as well as the sulfoxidation of the sulfur atom. These reactions are catalyzed by the sulfoxide synthase EgtB.

The apo structure of EgtB was phased using the anomalous signal of its intrinsic sulfur atoms (11 sulfurs from methionines and cysteines) in a recently developed data collection approach (Weinert *et al.*, 2015). EgtB consists of an N-terminal DinB domain (residues 7-150), formed by a four α-helical bundle containing the three Fe(II)-chelating residues H51, H134 and H138, and a C-terminal C-type lectin fold (CLec) domain, the latter poor of secondary structure elements. The two domains are connected by an extended two-stranded β-sheet linker formed by residues 151-210.

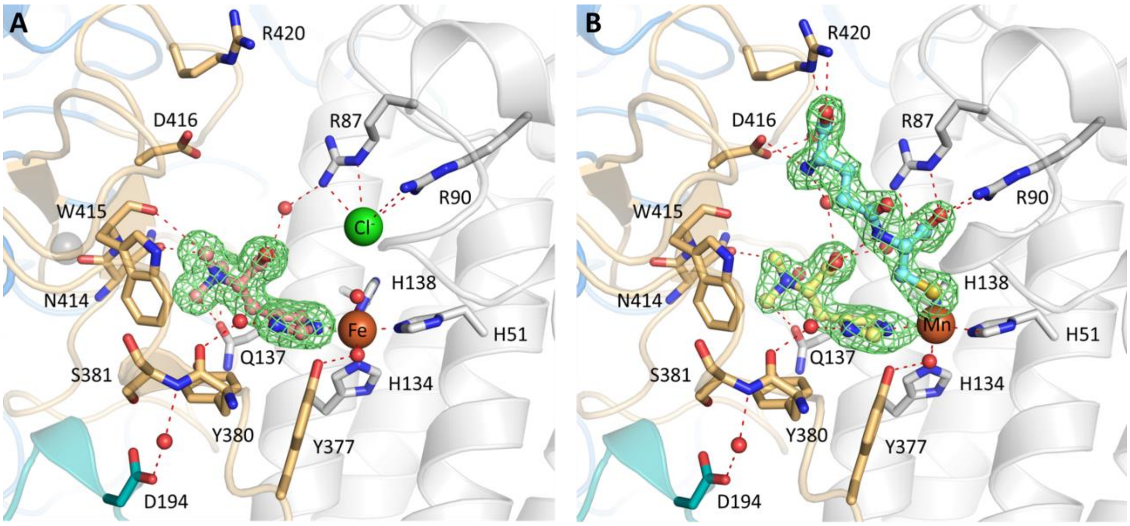


Figure 84: Active site of EgtB. Left: EgtB in complex with N^{α},N^{α},N^{α}-trimethyl-histidine (salmon). Unbiased $m|F_{obs}|-D|F_{calc}|$ density for TMH is shown in green. The Nπ of the imidazole moiety of TMH coordinates the Fe(II) ion together with H51, H134 and H138. Right: EgtB in complex with N^{α},N^{α}-dimethyl-histidine (yellow) and γ-glutamyl-cysteine (cyan). Unbiased $m|F_{obs}|-D|F_{calc}|$ density (σ-level = 2.5) for DMH and 3GC is shown in green. A dense array of hydrogen bonds positions the substrates in the active site of EgtB. 3GC is establishes hydrogen bonds to R87, R90 and R420, as well as D416 and the carboxyl group of DMH. The sulfur atom coordinated at the metal center. The orientation of the two substrates relative to each other gives hints for the stereochemical center of the reaction product.

The catalytic site of EgtB is located in a 15 Å deep and 10 Å wide pocket formed by residues 375-420 of the hypervariable region (Figure 84, orange) of the CLec domain and residues of helices α2 and α4 of the DinB domain of EgtB. In this study, it was possible to identify H51, H134 and H138 as the Fe(II)-complexing residues and direct ligands to the metal center, correcting prior bioinformatic analysis given in the literature. The fourth metal ligand is the Nπ of the imidazole moiety of N^{α},N^{α},N^{α}-trimethyl-histidine or N^{α},N^{α}-dimethyl-histidine (Figure 84A+B), the latter serving as a substrate analogue in the quaternary complex. The substrate is bound to EgtB by hydrogen bond interaction of the carboxyl groups with R87, R90 and R420. The sulfur atom of γ-glutamyl-cysteine is the fifth ligand coordinated at the metal center. The N^{α} amino function of γ-glutamyl-cysteine establishes a hydrogen bond to D416. Interestingly, γ-glutamyl-cysteine forms hydrogen bonds to the carboxyl group of N^{α},N^{α}-dimethyl-histidine. In turn, N^{α},N^{α}-dimethyl-histidine interacts with Q137 and the carbonyl group of Y380 via hydrogen bonds. The indole side chain of W415 is in hydrophobic contact with the methylated α-amino group of N^{α},N^{α}-dimethyl-histidine. Additionally, a tunnel leading from the surface of EgtB to the catalytic Fe(II) center was identified as the probable path taken by an oxygen molecule involved in the reaction (Goncharenko *et al.*, 2015).

On the basis of the structural and kinetic data, a catalytic mechanism for this novel class of sulfoxide synthases was proposed. This mechanism allowed the prediction of the sulfoxide stereochemistry of the product of the EgtB-catalyzed reaction to be in *S* configuration. The structure of EgtB and its mechanism are distinct from any known sulfoxide synthase.

13.3 Ntn Amidohydrolase EgtC

The Ntn hydrolase EgtC catalyzes the hydrolysis of γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide to glutamate and N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide. Its characteristic Ntn hydrolase (Pfam CL0052) (Brannigan *et al.*, 1995) nucleophile is provided by N-terminal cysteine residue C2. The apo of EgtC has been determined previously (Vit, 2010), while structures of EgtC with glutamine as well as of an inactive EgtC variant EgtC_{C2A} with its substrate γ-glutamyl-N^{α},N^{α},N^{α}-trimethyl-histidinyl cysteine sulfoxide were determined in this study.

EgtC forms tetramers in size exclusion chromatography and in the crystals (Figure 85A). Interestingly, dimerization of EgtC appears obligatory for activity because the substrate binding site is formed by residues from both molecules of the dimer (Figure 85A).

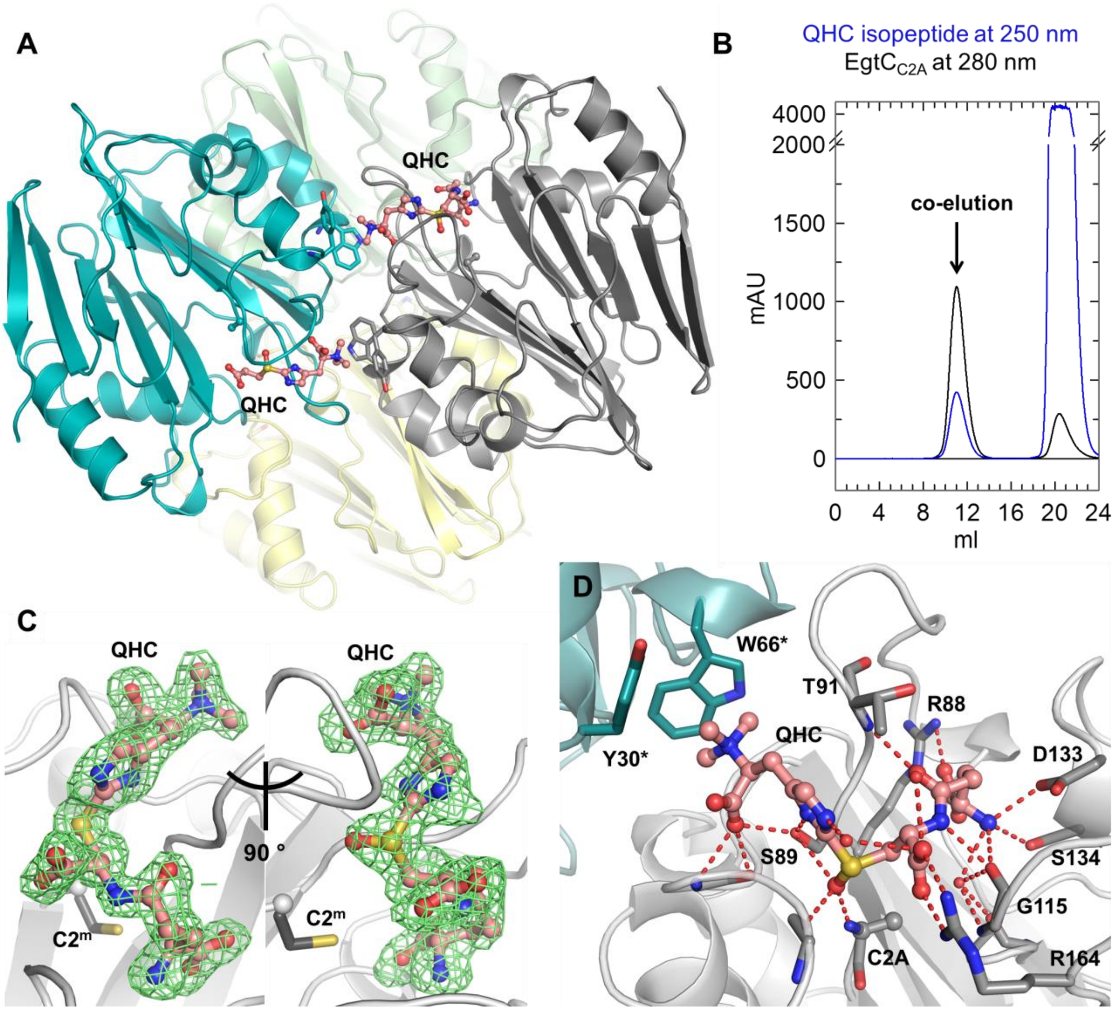


Figure 85: The crystal structure of EgtC$_{C2A}$ in complex with γ-glutamyl-N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide (QHC). A: EgtC is at least an obligate dimer (yellow/teal), because QHC (salmon) binding involves two residues from the second EgtC molecule. However, a tetramer is more likely because of the amount of buried surface. B: QHC coelutes with EgtC$_{C2A}$ from an S200 10/300 GL size exclusion chromatography column. C: The presence of QHC in the EgtC$_{C2A}$ active site is clearly indicated by positive m|F$_{obs}$|-D|F$_{calc}$| electron density (σ-level = 2). This density could be used to determine the stereochemistry of the sulfoxide moiety of QHC. C2^m is modeled from the wild type protein. D: QHC is recognized by a dense array of hydrogen bonds.

Positive m|F$_{obs}$|-D|F$_{calc}$| electron density clearly confirms the presence of the substrate in the EgtC binding pocket (Figure 85C), with the substrate stabilized by a dense network of hydrogen bonds (Figure 85D). The electron density was used to elucidate the stereochemistry of the sulfoxide moiety of γ-glutamyl-N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide. The sulfoxide moiety of the substrate does indeed show the previously predicted *S* configuration (Figure 85C, Goncharenko *et al.*, 2015). However, the mode of substrate binding to EgtC seems to be unproductive, because the C2 nucleophile is too far away from the carbonyl carbon at which the nucleophile should attack. Interestingly, a flexible loop A90-P95 is located above the active site, either adopting an open (EgtC with glutamine) or a closed (EgtC$_{C2A}$ with its substrate) conformation, thereby generation a more open or closed active site. This loop is probably able to push γ-glutamyl-N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidinyl cysteine sulfoxide into the active site of EgtC, leading to a more productive substrate binding mode in which the hydrolysis reaction can take place. This model is in agreement with the hypothesis that enzymes of this family undergo a conformational change upon substrate binding (Isupov *et al.*, 1996).

13.4 Inhibition of EgtD by Histidine Derivatives

Interestingly, EgtD is essential for ergothioneine biosynthesis in Mycobacteria (Cheah *et al.*, 2013). Inhibition of this pathway might present a target for novel therapeutics, for instance against *M. tuberculosis*. Thus, the structure of EgtD provides the basis for rational design of ergothioneine biosynthesis inhibitors (Vit *et al.*, 2015).

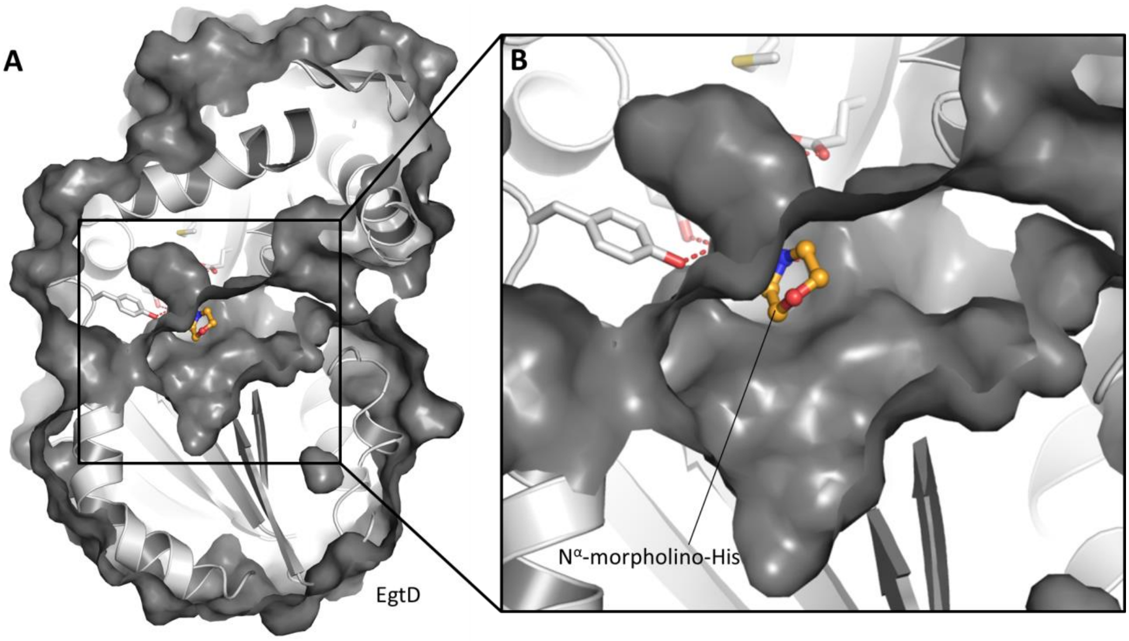


Figure 86: Future design of EgtD inhibitors. A: Overall structure of EgtD with N$^{\alpha}$-morpholino-histidine. B: Close-up of the N$^{\alpha}$-morpholino-histidine binding site. The N$^{\alpha}$-morpholino moiety, in boat conformation, points into the large cleft in EgtD, where SAM/SAH is normally bound. Possibly, EgtD is able to harbor inhibitors with more bulky α-amino group modifications.

Potential histidine derivative inhibitors with modification on the α-amino group of the compounds were designed by Florian Seebeck and coworkers. The compounds α-chloro-histidine, N$^{\alpha}$,N$^{\alpha}$,N$^{\alpha}$-trimethyl-histidine, N$^{\alpha}$-pyrrolidino-histidine and N$^{\alpha}$-morpholino-histidine were successfully assayed regarding their inhibitory potential towards EgtD, and their K_I's were determined as 2.5 µM, 16.1 µM, 27.2 µM and 22.0 µM, respectively (L. Misson, unpublished data). Hence, α-chloro-histidine is the most potent EgtD inhibitor. This is surprising, because it means that a positive charge on the α-amino group is not required for substrate recognition by EgtD. Crystal structures of EgtD with each of the four EgtD inhibitors were obtained by soaking approaches. The inhibitors are well-defined in the active site. Possibly, EgtD is even able to harbor inhibitors with more bulky α-amino group modifications (Figure 86) that reach into the SAM/SAH binding site.

13.5 Outlook – Future Studies on Ergothioneine-Biosynthetic Proteins

Further investigations on ergothioneine biosynthesis may include detailed mechanistic studies on EgtB, using enzyme engineering approaches to prove the proposed reaction mechanism, or to shift the substrate specificity of EgtB to other cysteine derivatives by generating tailor-made sulfoxide synthases.

EgtC been characterized structurally, but not in a kinetic assay, which would be a future task. In this assay, the activity of EgtC towards other compounds, for instance glutamine, should also be examined. Furthermore, EgtC was demonstrated to be an obligate dimer, because the substrate binding pocket is formed by residues from two distinct EgtC molecules. However, oligomer assembly predictions by the PISA server indicate that EgtC is a homotetramer (Krissinel & Henrick, 2007). In order gain additional and unambiguous information on the oligomeric state of EgtC, dynamic light scattering (DLS) or SAXS measurements of EgtC could be conducted in the future.

Inhibitor design for EgtD has been started, but requires a larger variety of compounds to improve the affinity of the inhibitors towards the enzyme. The inhibitors can be easily tested in the EgtD coupled assay established by L. Misson, and their K_I's can be determined. Crystal structures of the inhibitors in the active site of EgtD facilitate further improvement of the initial inhibitors.

Three of the five mycobacterial ergothioneine biosynthesis enzymes have successfully been characterized using structural biology and biophysical as well as biochemical approaches. In order to characterize the whole mycobacterial ergothioneine biosynthetic pathway, the two remaining enzymes, the glutamate-cysteine ligase EgtA and the PLP-dependent lyase EgtE should be subjected to structure determination. For EgtA, structure determination and kinetic data would complement already existing biochemical characterization of a similar enzyme from *Haloferax volcanii* (Malki *et al.*, 2009). Mycobacterial EgtE and its fungal variant Nc-Egt1 from *N. crassa* have not been structurally characterized at all. Hence, structure determination of EgtE would elucidate the last step in ergothioneine biosynthesis and complete the picture of the ergothioneine biosynthesis pathway.

14 Supporting Information

This section contains the supporting information for the publications presented in this study. The supporting information includes additional details experimental methods as well as further results and figures. Crystallographic data collection, processing and refinement statistics are also found in this section.

14.1 Supporting Information for Publication A

Full supporting information is available online:

http://onlinelibrary.wiley.com/store/10.1002/cbic.201402522/asset/supinfo/cbic_201402522_sm_ miscellaneous_information.pdf?v=1&s=fee4509a74cd3ce5fe81f309f30ba42b6f86947a

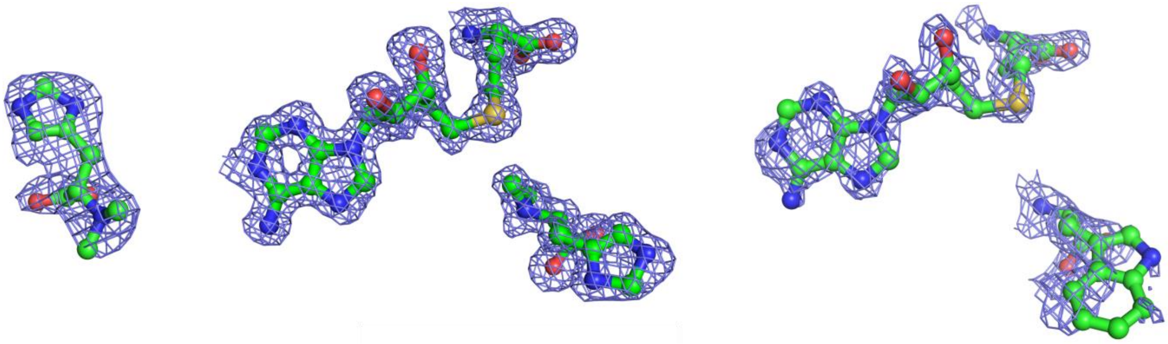


Figure S1: m|F_{obs}|-D|F_{calc}| densities. Left: Density for N^{α},N^{α}-dimethyl-histidine (2σ) from EgtD in complex with N^{α},N^{α}-dimethyl-histidine. Middle: Density for S-adenosyl-homocysteine (SAH) and N^{α},N^{α}-dimethyl-histidine (2σ) from EgtD in complex with SAH and N^{α},N^{α}-dimethyl-histidine. Right: Density for S-adenosyl-homocysteine and tryptophan (1.5σ) from EgtD$_{M252,E282}$ in complex with SAH and tryptophan.

Table S1: Data collection statistics for *M. smegmatis* EgtD apo structure and complexes. Values in parentheses are for the highest resolution shell. AVI=N^{α},N^{α}-dimethyl-L-histidine, SAH=S- adenosyl-L-homocysteine, TRP=L-tryptophan, HyoA = tryptophan-specific EgtD variant ($\text{EgtD}_{\text{M252V,E282A}}$).

Dataset	Se SAD[+] EgtD_apo	EgtD_Se_apo	EgtD_AVI	EgtD_AVI_SAH	HyoA_TRP_SAH
Wavelength (Å) / beamline[≠]	0.9786 / SLS, PXII	0.9786 / SLS, PXII	0.9184 / BESSY II, 14.1	0.9184 / BESSY II, 14.1	0.9184 / BESSY II, 14.1
Resolution range (Å)	48.72–2.77 (2.92 – 2.77)	48.72 –1.75 (1.78-1.75)	50-1.9 (2.0 – 1.9)	50-1.51 (1.59-1.51)	41.86-1.8 (1.9-1.8)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P2_1$
Unit cell parameters (Å)	71.8 75.5 138.7	71.8 75.5 138.7	72.0 74.8 137.8	56.7 67.5 79.7	79.04 67.63 112.13
(°)	90 90 90	90 90 90	90 90 90	90 110.98 90	90 109.24 90
Mosaicity (°)[†]	0.115	0.115	0.313	0.148	0.22
Total No. of measured reflections	250555 (39024)	1013484 (57635)	294015 (42418)	275925 (38056)	391092 (56192)
Unique reflections	19504 (2761)	74528 (3977)	59396 (8575)	87536 (12354)	103604 (15045)
Multiplicity	6.9 (7.4)	13.6 (14.5)	5 (4.9)	3.2 (3.1)	3.8 (3.7)
Mean I/σ(I)	43.6 (28.1)[§]	18.8 (2.5)	10.9 (2.3)	8.9 (2.0)	7.4 (2.1)
Completeness (%)	99.1 (98.6)[+]	97.4 (95.8)	100 (100)	98.8 (95.9)	100 (99.9)
R_{meas} (%)[‡]	5.0 (8.9)	11.0 (132.2)	10.5 (75.4)	10.6 (66.5)	12.8 (68.6)
$R_{p.i.m.}$ (%)[$]	1.9 (3.3)	3.0 (34.5)	4.7 (33.5)	5.9 (36.9)	6.5 (35.1)

[+] Data collection statistics for unmerged Friedel pairs.

[≠] SLS: Swiss Light Source (Paul Scherrer Institute, Villigen, CH), BESSY: Berlin Electron Storage Ring Society for Synchrotron Radiation (Helmholtz Zentrum Berlin, Berlin, D).

[†] Mosaicity values reported by *XDS*.[1]

[§] The anomalous signal is significant to 2.77 Å resolution according to *Xtriage* from *Phenix*.[2] Hence, the data were cut at mean I/σ(I) = 28.1 in the highest resolution shell.

[‡] $R_{meas} = \Sigma_{hkl} (N/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where N is the number of observations of the reflection with index hkl and I_i is the intensity of its i^{th} observation.

[$] $R_{p.i.m.} = \Sigma_{hkl} (1/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$.[3]

Table S2: Refinement statistics for *M. smegmatis* EgtD apo structure and complexes. Values in parentheses are for the highest resolution shell. AVI=N$^{\alpha}$,N$^{\alpha}$-dimethyl-L-histidine, SAH=S-adenosyl-L-homocysteine, TRP=L-tryptophan, HyoA = tryptophan-specific EgtD variant (EgtDM252V,E282A).

Dataset	EgtD_Se_apo	EgtD_AVI	EgtD_AVI_SAH	HyoA_TRP_SAH
Resolution range (Å)	48.72 –1.75	50-1.9	50-1.51	41.86-1.8
	(1.82-1.75)	(2.0 – 1.9)	(1.59-1.51)	(1.9-1.8)
R_{work} (%)	15.64 (22.9)	19.8 (41.6)	14.8 (24.1)	19.4 (27.8)
R_{free} (%)	18.66 (26.5)	24.1 (44.9)	18.0 (27.4)	23.9 (36.4)
No. of non-H atoms				
Protein	4999	4933	4987	9679
Ion	8	25	7	3
Ligand	-	26	78	164
Water	834	680	985	1371
R.m.s. deviations				
Bonds (Å)	0.009	0.003	0.008	0.007
Angles (°)	1.116	0.790	1.278	1.104
Average B factors (Å^2)				
Protein	25	29	13	22
Ion	34	33	30	23
Ligand	-	21	8	17
Water	39	35	27	30
Ramachandran plot				
Favored regions (%)	98.2	98.4	98.6	97.9
Outliers (%)	0.2	0	0	0.1
MolProbity score#	0.95	0.88	0.94	1.15
PDB entry code	4PIM	4PIN	4PIO	4PIP

#As reported by *MolProbity*.[4]

References

[1] W. Kabsch, *Acta Crystallog. D Biol. Crystallogr.* **2010**, *66*, 125-132.

[2] a) P. H. Zwart, R. W. Grosse-Kunstleve, P.D. Adams, CCP4 Newsletter, **2005**, 43. b) P. R. Evans, *Acta Crystallog. D Biol. Crystallogr.* **2011**, *67*, 282-292.

[3] M. S. Weiss, *J. Appl. Crystallogr*. **2001**, *34*, 130–135.

[4] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallog. D Biol. Crystallogr.* **2009**, *66*, 12-21

14.2 Supporting Information for Publication B

Full supporting information is available online:

http://onlinelibrary.wiley.com/store/10.1002/anie.201410045/asset/supinfo/anie_201410045_sm_miscellaneous_information.pdf?v=1&s=2883ace3ed2b1a75519a3845f7420232a2636e2b

Materials. All standard reagents were purchased from Aldrich/Sigma if not otherwise stated. Synthetic oligonucleotides were purchased from Microsynth, Switzerland. N-glutaryl-cysteine was synthesized from glutaric anhydride and cysteine and characterized by ESI-MS (m/z calc. 236.059; meas.: 236.1) and ^{1}H NMR (400 MHz, DMSO---d$_6$) δ ppm 7.43 (d, J = 6.9 Hz, 1H), 4.01 (dt, J = 6.8, 5.1 Hz, 1H), 2.82 (dd, J = 13.0, 4.7 Hz, 1H), 2.73 (dd, J = 13.0, 5.6 Hz, 1H), 2.13 (t, J = 7.4 Hz, 2H), 2.04 (t, J = 7.4 Hz, 2H), 1.67 (p, J = 7.4 Hz, 2H).

Recombinant EgtB constructs. Cloning and production of EgtB$_{smegmatis}$ was described previously.[1] The gene for EgtB from *Mycobacterium thermoresistibile* (EgtB$_{thermo}$, WP_003925249) was codon-optimized for protein production in *E. coli* and purchased from GenScript. The gene was ligated as NdeI-XhoI fragment into a pET28b cloning vector. The corresponding construct induced *E. coli* cells to produce the following proteins:

EgtB$_{smegmatis}$:

```
GSSHHHHHHSSGLVPRGSHMIARETLADELARARERTLRLVEFDDAELHRQYNPLMSPLVWDLAHIGQQ
EELWLLRDGNPDRPGMLAPEVDRLYDAFEHSRASRVNLPLLPPSDARAYCATVRAKALDTLDTLPEDDP
GFRFALVISHENQHDETMLQALNLREGPPLLDTGTPLPTGRPGVAGTSVLVPGGPFVLGVDALTEPHSL
DNERPAHVVDIPSFRIGRVPVTNAEWREFIDDGGYDEPRWWSPRGWAHRQEAGLVAPQFWNPDGTRTRF
GHIEEIPGDEPVQHVTFFEAEAYAAWAGARLPTEIEWEKACAWDPVAGARRRFPWGSAQPSAALANLGG
DARRPAPVGAYPAGASAYGAEQMLGDVWEWTSSPLRPWPGFTPMIYERYSTPFFEGTTSGDYRVLRGGS
WAVAPGILRPSFRNWDHPIRRQIFSGVRLAWDV
```

m/z(EgtB$_{smegmatis}$): calc.: 49859.76 Da, meas.: 49998.4 Da

ε$_{280}$(EgtB$_{smegmatis}$): 108400 M^{-1} cm^{-1}

EgtB$_{thermo}$:

```
GSSHHHHHHSSGLVPRGSHMGVAVPHRAELARQLIDARNRTLRLVDFDDAELRRQYDPLMSPLVWDLAH
IGQQEELWLLRGGDPRRPGLLEPAVEQLYDAFVHPRASRVHLPLLSPAQARRFCATVRSAVLDALDRLP
EDADTFAFGMVVSHEHQHDETMLQALNLRSGEPLLGSGTALPPGRPGVAGTSVLVPGGPFVLGVDLADE
PYALDNERPAHVVDVPAFRIGRVPVTNAEWRAFIDDGGYRQRRWWSDAGWAYRCEAGLTAPQFWNPDGT
RTRFGHVEDIPPDEPVQHVTYFEAEAYAAWAGARLPTEIEWEKACAWDPATGRRRRYPWGDAAPTAALA
NLGGDALRPAPVGAYPAGASACGAEQMLGDVWEWTSSPLRPWPGFTPMIYQRYSQPFFEGAGSGDYRVL
RGGSWAVAADILRPSFRNWDHPIRRQIFAGVRLAWDVDRQTARPGPVGGC
```

m/z(EgtB$_{thermo}$): calc.: 51342.67 Da, meas. 51373.70 Da; Delta: 31 Da*

ε280(EgtB$_{thermo}$): 111380 M^{-1} cm^{-1}

* The calculated mass is 31 Da smaller that the measured mass. This may be due to two oxidation events, for example on methionine and cysteine residues. The exposed C-terminal Cys residue (Cys446) is a possible recipient of this oxidation. However, but this residue is invisible in the X-ray structures.

EgtB variants. EgtB$_{smegmatis,C100S}$, EgtB$_{D194N}$ and EgtB$_{D416N}$ were constructed by primer extension using the following primers. The resulting fragments were cloned into pET28b vectors. For protein crystallization we cloned the EgtB$_{thermo}$ gene into a modified vector pET19m to encode an EgtB fusion construct with an N-terminal His$_6$-tag followed by a TEV (tobacco etch virus) protease cleavage site.

D416Ns: 5‛-TAT ATC GCA ATT GGA ACC ACC CGA TTC -3‛
D416Na: 5′-TAT AGA ATC GGG TGG TTC CAA TTG CGA -3′
EgtB$_{thermo}$s: 5′-TAT ACA TAT GGG TGT CGC CGT GC -3′
EgtD$_{thermo}$a: 5′-ATA TCT CCG AGC TAA CAA CCA CCC ACC GG -3′
C100Ss: 5‛-ATG CGC GCG CCT ACT CTG CGA CGG TGC GGG CCA A-3‛
C100Sa: 5′-CCG CAC CGT CGC AGA GTA GGC GCG CGC ATC CGA A-3′
EgtB$_{smegmatis}$s: 5′-ATA TCA TAT GAT CGC ACG CGA GAC ACT-3′
EgtD$_{smegmatis}$a: 5′-ATA TCT CGA GTT AGA CGT CCC AGG CCA G-3′

m/z(EgtBD416N): calc.: 51341.68 Da, meas.: 51373 Da
ε$_{280}$(EgtBD416N): 111380 M^{-1} cm^{-1}

m/z(EgtB*smegmatis*,C100S): calc.: 49843.7 Da, meas.: 49843 Da;
ε$_{280}$(EgtB*smegmatis,*C100S): 108400 M^{-1} cm^{-1}

m/z(EgtB*thermo* without His$_{6}$-tag): calc.: 49504.73 Da, meas.: 49536 Da; Delta: 31 Da

Recombinant protein production. *E. coli* BL21(DE3) transformed with the EgtB expression vectors were cultured in LB medium supplemented with 50 µg/mL kanamycin (pET28b) or 100 µg/mL ampicillin (pET19m) and 30 µg/mL chloramphenicol. At an optical density of OD$_{600}$ of 0.8, protein production was induced with 0.5 mM IPTG for 16 hours at 19 °C. Cells were harvested by centrifugation, resuspended in buffer A (300 mM NaCl, 50 mM Na_2HPO_4 pH 8.0) and lysed by sonication. After centrifugation for 1 h at 48000 rpm, the protein was purified using Ni^{2+} NTA Agarose beads (Qiagen) by washing the beads with buffer A containing 10 mM imidazole and then by elution in buffer A supplemented with 250 mM imidazole. The proteins were dialyzed into 50 mM Tris/HCl, 50 mM NaCl buffer, pH 8.0 and stored at -80°C. Protein homogeneity was assessed by SDS PAGE, protein concentration was determined by UV using calculated extinction coefficients.[2] Removal of the N-terminal histidine-tag from EgtB_{thermo} with recombinant TEV protease proceeded after Ni^{2+} NTA Agarose purification during dialysis against 50 mM Tris/HCl, 50 mM NaCl pH 7.5 overnight at 277 K. Unprocessed EgtB and TEV were separated by filtration through Ni^{2+} NTA Agarose. The protein was further purified by size exclusion chromatography (S300) and concentrated to 15 mg ml^{-1} and stored at -80°C.

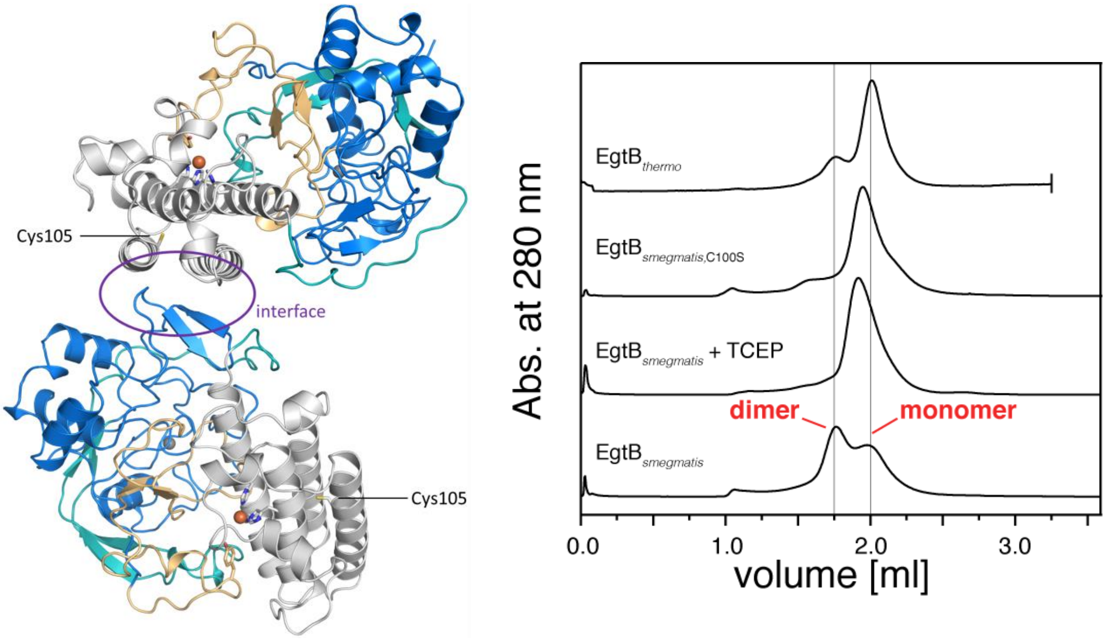


Figure S1: Quaternary structure of EgtBs. Left: Cartoon diagram of the asymmetric unit of EgtB_{thermo} indicating Cys105 and the crystallographic interface. The interface in EgtB_{thermo} does not form a disulfide bond because the equivalent Cys residues (Cys105) are not in close distance and point to the inside of the protein. Hence, EgtB_{thermo} is monomeric. In $\text{EgtB}_{smegmatis}$, the orthologous residue is Cys100, suggesting that this protein adopts a different dimer structure in solution. Right: Analytical size-exclusion chromatography of $\text{EgtB}_{smegmatis}$, $\text{EgtB}_{smegmatis,C100S}$ and EgtB_{thermo} (Superdex200 size exclusion chromatography, at room temperature using 50 mM Tris-HCl pH 7.4, 200 mM NaCl, as running buffer, flow rate: 0.2 mL/min). Purified $\text{EgtB}_{smegmatis}$ elutes as a monomer-dimer mixture. Treatment with TCEP turns the protein into a homogeneous monomer, suggesting that the dimer-interaction is mediated by an intramolecular disulfide bond. Consistently, the $\text{EgtB}_{smegmatis,C100S}$ variant does not dimerize during protein purification. The crystallographic dimer interface in EgtB_{thermo} does not form a disulfide bond and the equivalent Cys residues (Cys105) are not in close distance suggesting that the protein adopts a different dimer structure in solution. In the crystal structure of EgtB_{thermo} we also detected an intermolecular disulfide bond between two Cys243 residues.

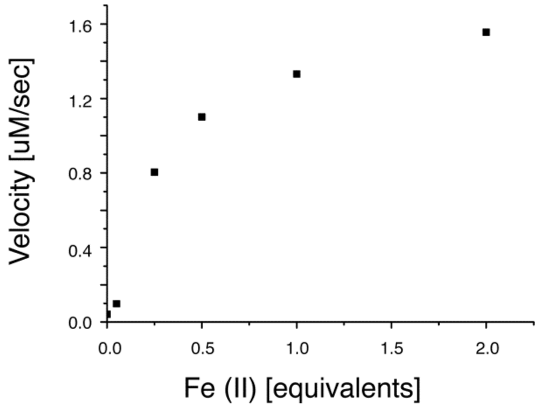
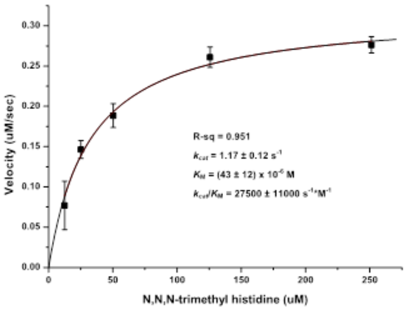


Figure S2: EgtB activity depends on the presence of 1 equivalent iron. EDTA dialyzed EgtB$_{smegmatis}$ showed no measurable activity when assayed in a reaction containing 100 mM HEPES, pH 8.0, 100 mM NaCl, 2 mM TCEP, 2 mM ascorbate, 0.4 µM TMH, 1.2 mM γGC and 1.6 µM EgtB$_{smegmatis}$ but no $FeSO_4$. Titration of this reaction with $FeSO_4$ induced 80% activity in the presence of 1 equivalent of iron (II) and full activity (V_{max}) in presence of > 2 equivalents of iron (II). Consequently all EgtB activities were determined in presence of 4 equivalents of $FeSO_4$. Comparison of EgtB activities in absence or presence of exogenous $FeSO_4$ suggest that the recombinant proteins purify to a significant extent as holoenzymes (Table S1). Iron content of purified EgtB proteins was also determined by a ferrozine-based colorimetric assay. The data displayed in Table S1 are averages from three independent measurements.

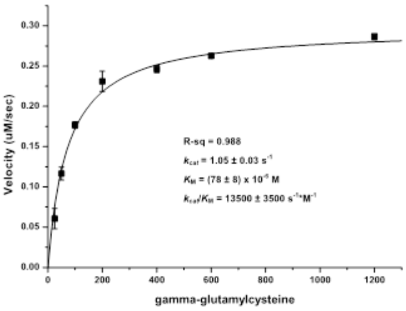
Table S1: Iron content in recombinant EgtB enzymes.

Protein	Ferrozine Assay (equivalent x 100)	Activity assay (% residual activity)
EgtB$_{thermo}$	98 ± 2	95 ± 2
EgtB$_{thermos,D416N}$	55 ± 6	61 ± 3
EgtB$_{smegmatis}$	61 ± 6	54 ± 4

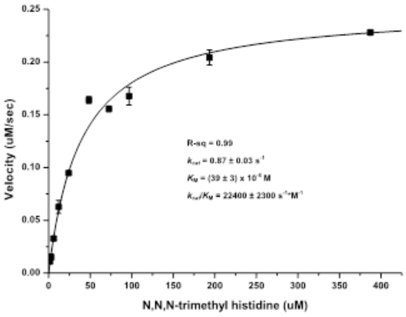
Figure S3 (below). Michaelis-Menten plots/Enzyme assay. All enzymes were analyzed under the following standard conditions. Reactions contained 100 mM HEPES, 100 mM NaCl, 2 mM TCEP, 2 mM ascorbate, sulfur acceptor TMH (or alternative sulfur acceptors) and EgtB. Reactions were started by addition of γGC and incubated at 26 °C. Aliquots of the reactions were quenched by addition of 20 µl 1 M phosphoric acid and analyzed by cation exchange HPLC using 20 mM phosphoric acid at pH 2 as a mobile phase.[3] Compounds were eluted in a NaCl gradient. All HPLC chromatograms were recorded at 265 nm. The data were fitted to the function $v = V_{max}[s]/(K_M + [s])$. All reported data represent averages of at least two independent measurements. Michaelis-Menten plots are shown below. K_{cat} and k_{cat}/K_M were determined in the presence of co-substrate at a concentration at least 3-fold higher than the corresponding K_M, cosubstrate and in air saturated buffers.



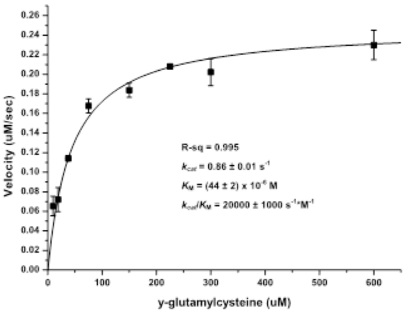
EgtB$_{smegmatis}$, [γGC] = 1200 μM



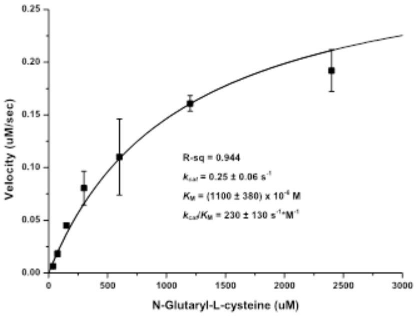
EgtB$_{smegmatis}$, [TMH] = 500 μM



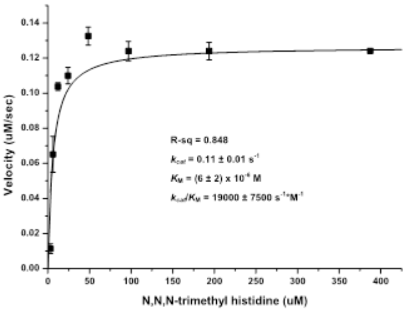
EgtB$_{thermo}$, [γGC] = 1200 μM



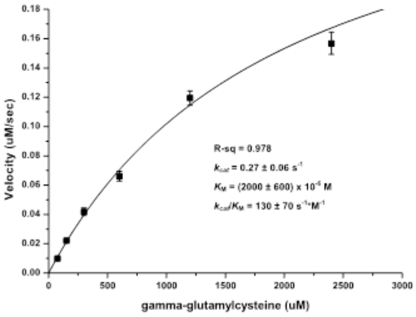
EgtB$_{thermo}$, [TMH] = 290 μM



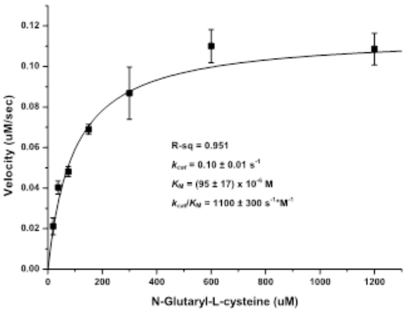
EgtB$_{thermo}$, [TMH] = 200 μM



EgtB$_{D416N}$, [γGC] = 1200 μM



EgtB$_{D416N}$, [TMH] = 200 μM



EgtB$_{D416N,}$ [TMH] = 200 μM

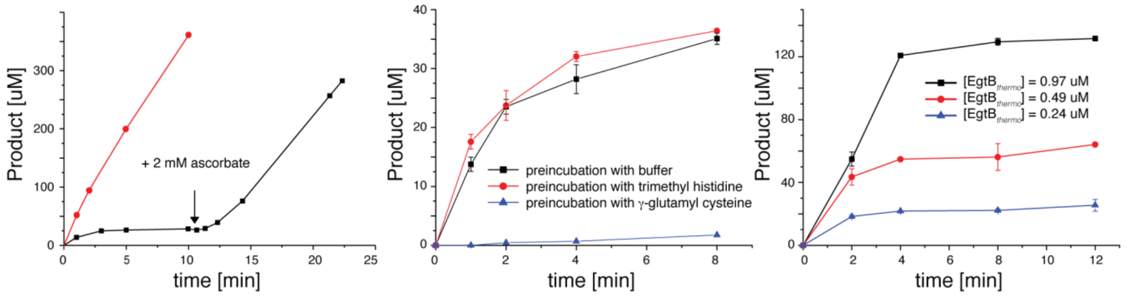


Figure S4. Left: γGC-dependent inactivation is reversible by addition of ascorbate. Product formation by EgtB$_{thermo}$ in a HEPES buffered solution (100 mM, pH 8.0), containing 100 mM NaCl, 0.8 mM TMH, 1.2 mM γGC and 0.9 µM EgtB$_{thermo}$ at 26 °C, was monitored by HPLC. On reaction contained 2 mM ascorbate (red line). In absence of ascorbate EgtB$_{thermo}$ activity ceased after 30 turnovers (black line). The enzyme reaction resumed after addition of 2 mM ascorbate. Middle: EgtB$_{thermo}$ is inactivated by preincubation with γGC. EgtB$_{thermo}$ was incubated at 4 °C for 1 h with either 0.5 mM γGC or 0.5 mM TMH or buffer as a control. The remaining activities of these proteins (1.4 µM) were assayed in reactions containing 100 mM HEPES buffer, pH 8.0, 100 mM NaCl, 2 mM TCEP, 0.8 mM TMH and 1.2 mM γGC. Product formation was monitored by HPLC. The initial rates of the three reactions indicate that incubation with γGC reduces EgtB$_{thermo}$ activity by at least 80-fold. Incubation with TMH leaves EgtB$_{thermo}$ activity unchanged. Right: Number of catalyzed turnovers by EgtB$_{thermo}$ in absence of ascorbate: Reaction mixtures containing HEPES buffer pH 8.0, 100 mM NaCl, 2 mM ascorbate, 0.4 mM TMH, 1.2 mM γGC, 2 mM TCEP were initiated by addition of 0.2, 0.5 or 1 µM EgtB$_{themo}$. Product formation was monitored by HPLC. Product formation ceases after 120 ± 20 turnovers.

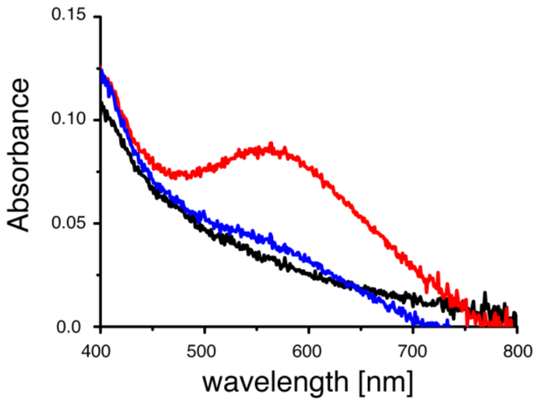


Figure S5. Absorption spectra of a) 100 µM EgtB$_{thermo}$ in 50 mM Tris (pH 8.0), 50 mM NaCl (black); b) sample a plus 500 µM γGC (red); c) sample b plus 1 mM ascorbate (blue). In the presence of γGC EgtB$_{thermo}$ is characterized by a strong absorbance band at 565 nm consistent with a charge transfer from γGC to the iron (III) center (LMCT – band).[4]

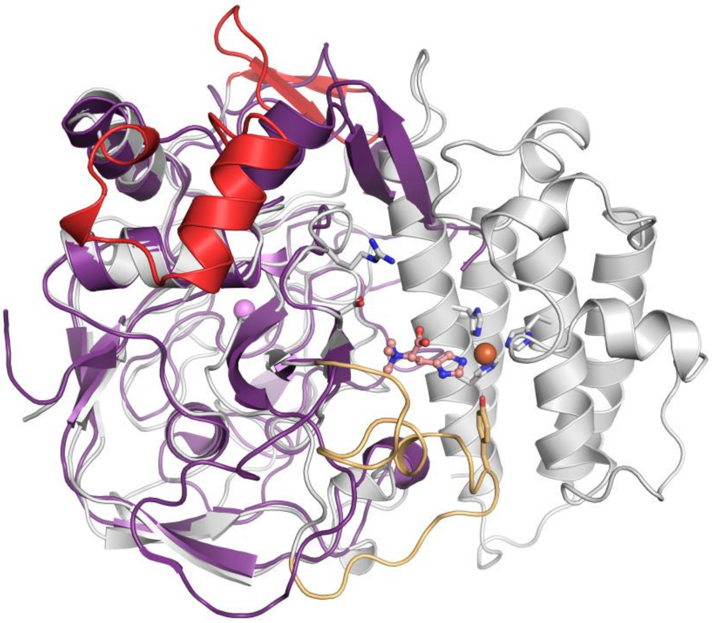


Figure S6. Superposition of EgtB$_{thermo}$ (grey/red/orange) and FGE$_{human}$ (2AIJ, violet).[5] The segments 227-269 (red) and 370-392 (orange) (EgtB$_{thermo}$ numbering) diverge significantly from the same segments in FGE$_{human}$ (violet). This unusual loop-rich structure is remarkably conserved in at least two proteins with entirely different functions and less than 30% sequence homology.[5-6] The backbones of EgtB$_{thermo}$ (residues 155-434) and FGE superimpose with an r.m.s.d. of 0.741 when two peripheral loops are excluded (residues 227-269 and 370 - 392). FGE uses a concave patch formed by a C-terminal segment of 30 residues (327-369, FGE$_{human}$) to harbor an active site. The same surface in EgtB$_{thermo}$ combines with the N-terminal DinB domain for form the sulfoxide synthase active site.

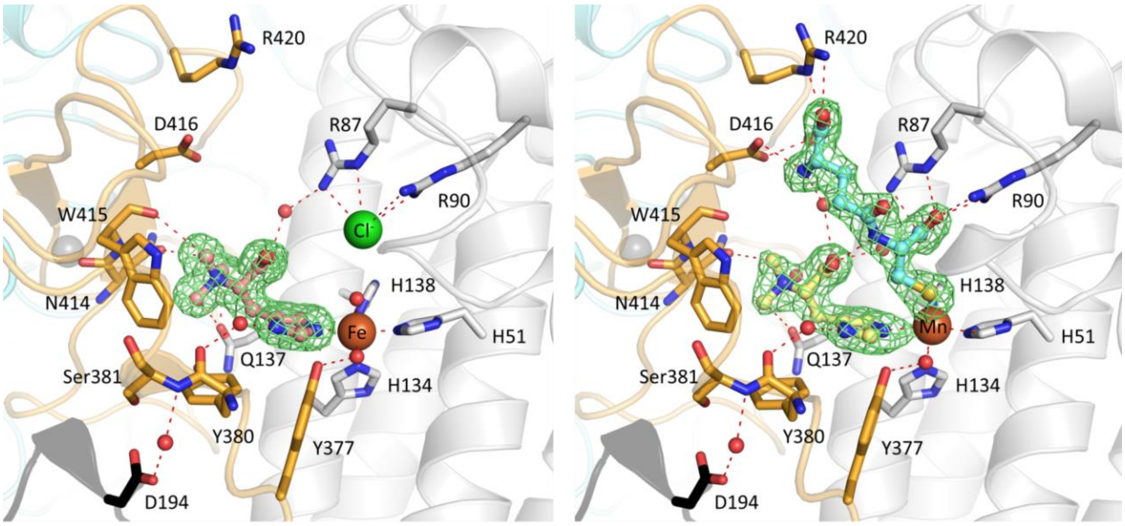


Figure S7: m|F$_{obs}$|-D|F$_{calc}$| electron densities (σ-level = 2.5) of bound ligands are shown. Left: Difference electron density for TMH bound to EgtB$_{thermo}$. Right: Difference electron density for DMH and γGC bound to EgtB$_{thermo}$.

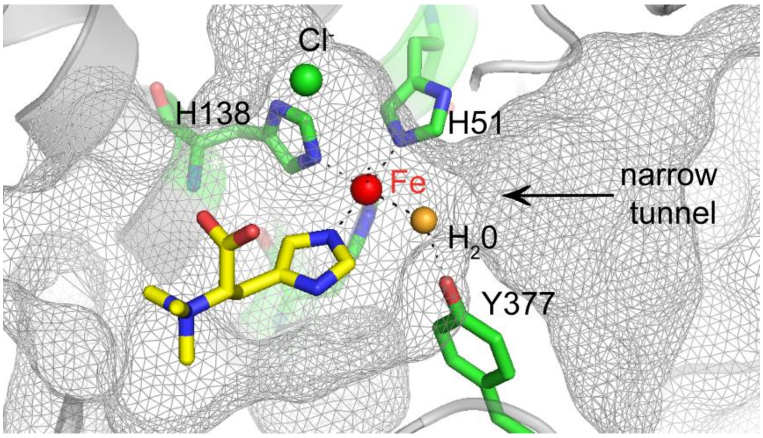


Figure S8: The view down the broad active site tunnel (15 Å deep and 10 Å wide) of the ternary complex reveals the iron binding site (active site residues in green) with ligated TMH (yellow). A water molecule (orange) ligates to iron and also hydrogen bonds with Tyr377. This ligand sits at the mouth of a narrow tunnel (arrow) which leads out to the protein exterior. It seems possible that this tunnel allows O_2 to reach the iron center to initiate the oxidation reaction.

Crystallization. Initial crystallization conditions of EgtB*thermo* were determined with the vapor diffusion method in a sitting drop 96-well format. Drops were set up using a dispensing robot (Honeybee, Zinsser Analytic) mixing 0.2 µL of EgtB solution (23 mg mL^{-1}) with 0.2 µL of reservoir solution equilibrated against 70 µL reservoir solution. The screens were stored at 20 °C in an automated imaging system (Formulatrix). Several initial hits were identified and optimized in 24-well plates using a hanging drop setup. 1 µL of protein solution (20 mg mL^{-1}) was mixed with 1 µL mother liquor and equilibrated against 500 µL of reservoir at 20 °C. The final condition of EgtB crystals contained 0.2 M $MgCl_2$, 0.1 M Tris/HCl pH 7 – 8, 6 - 10% (w/v) PEG 6000. Apo crystals of EgtB were cryoprotected in reservoir solution supplemented with 20% (v/v) glycerol, flash cooled in liquid nitrogen and tested on an X-ray home source including a Saturn 944+ CCD detector (Rigaku). Crystals of EgtB in complex with DMH or TMH were obtained by overnight soaking in reservoir solution supplemented with 15 % (v/v) glycerol and 10 mM ligand. EgtB crystals in complex with DMH and γQC were obtained by iterative addition of compounds to the reservoir solution (i) 80 mM manganese chloride, 100 mM glutamate and overnight soaking (ii) 10 mM DMH and three hours soaking and (iii) 10 mM γQC and 15 % (w/v) glycerol and soaking for two hours. Well-diffracting crystals of EgtB were sent to a synchrotron for data collection.

Data collection, data processing, structure solution and refinement. Single anomalous diffraction (SAD) data of an EgtB apo crystal were collected at a wavelength of 2.06642 Å on beamline X06DA at the Swiss Light Source (SLS, Paul Scherrer Institute, Villigen, Switzerland) using a Pilatus 2M detector (Dectris). A native dataset was collected from a different EgtB apo crystal at a wavelength of 1.0 Å. All data were indexed and integrated with *XDS*[7] and scaled with *AIMLESS*[8] from the *CCP4* program suite.[9] For the SAD data, hkl2map[10] was used to employ *SHELXC*[11] in extraction of the anomalous signal and *SHELXD*[12] in locating the anomalous scatterers, before *SHELXE* was used for density modification and automated model building. Since native data extended to a resolution of 1.7 Å, SAD data and native data were combined. An initial model was generated using *arp/Warp* 7.0.[13] Diffraction data of EgtB ligand complexes were collected at the SLS (Villigen, Switzerland) or at the BESSY II Synchrotron (HZB, Berlin, Germany) at cryogenic temperatures. Data were processed with *XDS* and *AIMLESS*. In order to solve the EgtB ligand complexes, the apo structure of EgtB_{thermo} was used for rigid body refinement in *REFMAC5*.[14] Manual adjustment of all models was done in *COOT*[15] which included the placement of ions and ligands. Ligands were prepared using *PRODRG*[16] and *eLBOW*[17] for refinement with phenix.refine.[18] The structures were validated with *MolProbity*.[19] Figures were prepared with *PyMOL*.[20] Data collection and refinement statistics are presented in Tables S2 and S3.

Table S2: Data collection statistics for *M. thermoresistibile* EgtB apo structure and complexes. Values in parentheses are for the highest resolution shell. AVI = N^{α},N^{α}-dimethyl-histidine, AVJ = N^{α},N^{α},N^{α}-trimethyl-histidine, 3GC = γ-glutamyl-cysteine.

Dataset	S SAD[+] EgtB_apo	EgtB_apo	EgtB_AVI	EgtB_AVJ	EgtB_AVI_3GC
Wavelength (Å) / beamline[≠]	2.06642 / SLS, PXIII	1.0 / SLS, PXIII	1.0 / SLS, PXIII	1.0 / SLS, PXIII	0.9184 / BESSY II, BL14.1
Resolution range (Å)	48.83-2.27 (2.33-2.27)	48.75-1.7 (1.73-1.7)	48.7 -1.85 (1.88-1.85)	48.97-1.6 (1.63-1.6)	47.83-1.98 (2.01-1.98)
Space group	$P4_32_12$ (96)	$P4_32_12$ (96)	$P4_32_12$ (96)	$P4_32_12$ (96)	$P4_32_12$ (96)
Unit cell parameters (Å) (°)	135 135 141.5 90 90 90	134.8 134.8 141.2 90 90 90	134.3 134.3 141.4 90 90 90	135.2 135.2 142 90 90 90	135.3 135.3 141.3 90 90 90
Mosaicity (°)[†]	0.078	0.113	0.157	0.094	0.072
Total No. of measured reflections	2559724 (68788)	3781286 (191591)	2904833 (147156)	4581822 (223195)	1216498 (56796)
Unique reflections	58661 (3294)	142371 (6966)	110295 (5375)	172557 (8463)	91436 (4456)
Multiplicity	43.6 (20.9)	26.6 (27.5)	26.3 (27.4)	26.6 (26.4)	13.3 (12.7)
Mean I/σ(I)	64.2 (10.5)	24.6 (2.7)	20.9 (1.9)	32 (2.3)	13.4 (2.1)
Completeness (%)	96.4 (71)	100 (100)	100 (100)	100 (100)	100 (100)
R_{meas} (%)[‡]	5.8 (30.2)	12.0 (160.8)	13.6 (240.2)	8.1 (187.1)	15.5 (132)
$R_{p.i.m.}$ (%)[\$]	1.2 (8.8)	2.3 (30.6)	2.6 (45.7)	1.6 (36.3)	4.2 (36.8)

[+] Data collection statistics for unmerged Friedel pairs.

[≠] SLS: Swiss Light Source (Paul Scherrer Institute, Villigen, CH), BESSY: Berlin Electron Storage Ring Society for Synchrotron Radiation (Helmholtz Zentrum Berlin, D)

[†] Mosaicity values reported by *XDS*.[7]

[§] The recommended resolution cut off for anomalous substructure search is 3.3 Å resolution (mean I/σ(I) =89.3) according to *Xtriage* from *Phenix*.[18] The data were cut at 2.27 Å resolution (mean I/σ(I) =10.5) in the highest resolution shell.

[‡] $R_{meas} = \Sigma_{hkl} (N/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where N is the number of observations of the reflection with index hkl and I_i is the intensity of its i^{th} observation.

[\$] $R_{p.i.m.} = \Sigma_{hkl} (1/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$.

Table S3: Refinement statistics for *M. thermoresistibile* EgtB apo structure and complexes. Values in parentheses are for the highest resolution shell. AVI = N^{α},N^{α}-dimethyl-histidine, AVJ = N^{α},N^{α},N^{α}-trimethyl-histidine, 3GC = γ-glutamyl-cysteine

Dataset	EgtB_apo	EgtB_AVI	EgtB_AVJ	EgtB_AVI_3GC
Resolution range (Å)	48.75-1.7	48.7 -1.85	48.97-1.6	47.83-1.98
	(1.73-1.7)	(1.88-1.85)	(1.63-1.6)	(2.01-1.98)
R_{work} (%)	13.8 (20.1)	15.9 (26.2)	14.5 (23.8)	15.1 (22.8)
R_{free} (%)	15.7 (21.8)	17.9 (27.8)	16.3 (28.5)	17.3 (25.8)
No. of non-H atoms				
Protein	6782	6708	6809	6827
Ion	8	14	13	23
Ligand	60	116	100	64
Water	1203	640	1106	1015
R.m.s. deviations				
Bonds (Å)	0.009	0.013	0.010	0.006
Angles (°)	1.214	1.360	1.307	1.045
Average B factors (Å^2)				
Protein	26	37	30	33
Ion	27	49	40	28
Ligand	38	51	28	28
Water	40	44	42	43
Ramachandran plot				
Favored regions (%)	99.0	98.9	99.1	98.8
Outliers (%)	0.2	0.2	0.2	0.2
MolProbity score#	0.95	1.08	1.04	0.83
PDB entry code	4X8B	-	4X8D	4X8E

#As reported by *MolProbity*.[19]

References

[1] F. P. Seebeck, *J. Am. Chem. Soc.* **2010**, *132*, 6632-6633.

[2] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, *Prot. Sci.* **1995**, *4*, 2411-2423.

[3] A. Braunshausen, F. P. Seebeck, *J. Am. Chem. Soc.* **2011**, *133*, 1757-1759.

[4] J. D. Gardner, B. S. Pierce, B. G. Fox, T. C. Brunold, *Biochemistry* **2010**, *49*, 6033-6041.

[5] T. Dierks, A. Dickmanns, A. Preusser-Kunze, B. Schmidt, M. Mariappan, K. von Figura, R. Ficner, M. G. Rudolph, *Cell* **2005**, *121*, 541-552.

[6] J. Le Coq, P. Ghosh, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14649-14653.

[7] W. Kabsch, *Acta Crystallog. D Biol. Crystallogr.* **2010**, *66*, 125-132.

[8] P. R. Evans, *Acta Crystallog. D Biol. Crystallogr.* **2011**, *67*, 282-292.

[9] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, Evans, P. R., R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, *et al.*, *Acta Crystallog. D Biol. Crystallogr.* **2011**, *67*.

[10] T. Pape, T. R. Schneider, *J. Appl. Cryst.*, 843-844.

[11] G. M. Sheldrick, *Acta Crystallog. D Biol. Crystallogr.* **2010**, *66*, 479-485.

[12] T. R. Schneider, G. M. Sheldrick, *Acta Crystallog. D Biol. Crystallogr.* **2002**, *58*, 1772-1779.

[13] G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, *Nat. Protoc.* **2008**, *3*, 1171-1179.

[14] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallog. D Biol. Crystallogr.* **2011**, *D67*, 355-367.

[15] P. Emsley, K. D. Cowtan, *Acta Crystallog. D Biol. Crystallogr.* **2004**, *60*, 2126-2132.

[16] A. W. Schuttelkopf, D. M. F. van Aalten, *Acta Crystallog. D Biol. Crystallogr.* **2004**, *60*, 1355-1363.

[17] N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, *Acta Crystallog. D Biol. Crystallogr.* **2009**, *65*, 1074-1080.

[18] P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, *Acta Crystallog. D Biol. Crystallogr.* **2012**, *68*, 352-367.

[19] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallog. D Biol. Crystallogr.* **2009**, *66*, 12-21.

[20] W. L. DeLano, *The PyMOL User's Manual*, DeLano Scientific, San Carlos, CA, USA; **2002**.

15 References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L.-W., Jain, S., Kapral, G. J., Grosse Kunstleve, R. W., et al. (2011). *Methods*. **55**, 94–106.

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367.

Ahuja, E. G., Janning, P., Mentel, M., Graebsch, A., Breinbauer, R., Hiller, W., Costisella, B., Thomashow, L. S., Mavrodi, D. V., & Blankenfeldt, W. (2008). *J. Am. Chem. Soc.* **130**, 17053–17061.

Akanmu, D., Cecchini, R., Aruoma, O. I., & Halliwell, B. (1991). *Arch. Biochem. Biophys.* **288**, 10–16.

De Angelis, J., Gastel, J., Klein, D. C., & Cole, P. A. (1998). *J. Biol. Chem.* **273**, 3045–3050.

Armstrong, R. N. (2000). *Biochemistry*. **39**, 13625–13632.

Askari, A. & Melville, D. B. (1962). *J. Biol. Chem.* **237**, 1615–1618.

Auld, D. S. (2001). *Biometals Int. J. Role Met. Ions Biol. Biochem. Med.* **14**, 271–313.

Baker, Y. R., Galloway, W. R. J. D., Hodgkinson, J. T., & Spring, D. R. (2013). *Molecules*. **18**, 11783–11796.

Baron, S. S. & Rowe, J. J. (1981). *Antimicrob. Agents Chemother.* **20**, 814–820.

Bartlett, G. J., Porter, C. T., Borkakoti, N., & Thornton, J. M. (2002). *J. Mol. Biol.* **324**, 105–121.

Bello, M. H., Barrera-Perez, V., Morin, D., & Epstein, L. (2012). *Fungal Genet. Biol.* **49**, 160–172.

Bello, M. H., Mogannam, J. C., Morin, D., & Epstein, L. (2014). *Fungal Genet. Biol.* **73**, 120–127.

Bergdoll, M., Eltis, L. D., Cameron, A. D., Dumas, P., & Bolin, J. T. (1998). *Protein Sci.* **7**, 1661–1670.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bittner, J. (2012). B. Sc. Thesis: Charakterisierung von PA3127 und PA3956 aus *Pseudomonas aeruginosa*. Universität Bayreuth.

Blankenfeldt, W., Kuzin, A. P., Skarina, T., Korniyenko, Y., Tong, L., Bayer, P., Janning, P., Thomashow, L. S., & Mavrodi, D. V. (2004). *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16431–16436.

Blankenfeldt, W. & Parsons, J. F. (2014). *Curr. Opin. Struct. Biol.* **29**, 26–33.

Botta, C., Di Giorgio, C., Sabatier, A.-S., & De Méo, M. (2008). *J. Photochem. Photobiol. B*. **91**, 24–34.

Brannigan, J. A., Dodson, G., Duggleby, H. J., Moody, P. C. E., Smith, J. L., Tomchick, D. R., & Murzin, A. G. (1995). *Nature*. **378**, 416–419.

Caldwell, C. C., Chen, Y., Goetzmann, H. S., Hao, Y., Borchers, M. T., Hassett, D. J., Young, L. R., Mavrodi, D., Thomashow, L., & Lau, G. W. (2009). *Am. J. Pathol.* **175**, 2473–2488.

Cameron, A. D., Olin, B., Ridderström, M., Mannervik, B., & Jones, T. A. (1997). *EMBO J.* **16**, 3386–3395.

Carollo, C. A., Calil, A. L. A., Schiave, L. A., Guaratini, T., Roberts, D. W., Lopes, N. P., & Braga, G. U. L. (2010). *Fungal Biol.* **114**, 473–480.

Cheah, I. K. & Halliwell, B. (2012). *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1822**, 784–793.

Cheah, I. K., Ong, R. L. S., Gruber, J., Yew, T. S. K., Ng, L. F., Chen, C. B., & Halliwell, B. (2013). *Free Radic. Res.* **47**, 1036–1045.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2009). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21.

DeLano, W. L. (2002). *DeLano Sci.* San Carlos, CA, USA.

Denning, G. M., Railsback, M. A., Rasmussen, G. T., Cox, C. D., & Britigan, B. E. (1998a). *Am. J. Physiol.* **274**, L893–L900.

Denning, G. M., Wollenweber, L. A., Railsback, M. A., Cox, C. D., Stoll, L. L., & Britigan, B. E. (1998b). *Infect. Immun.* **66**, 5777–5784.

Diederichs, K. & Karplus, P. A. (1997). *Nat. Struct. Mol. Biol.* **4**, 269–275.

Dietrich, L. E. P., Price-Whelan, A., Petersen, A., Whiteley, M., & Newman, D. K. (2006). *Mol. Microbiol.* **61**, 1308–1321.

Dietrich, L. E. P., Teal, T. K., Price-Whelan, A., & Newman, D. K. (2008). *Science*. **321**, 1203–1206.

Dorgan, K. M., Wooderchak, W. L., Wynn, D. P., Karschner, E. L., Alfaro, J. F., Cui, Y., Zhou, Z. S., & Hevel, J. M. (2006). *Anal. Biochem.* **350**, 249–255.

Emerson, J., Rosenfeld, M., McNamara, S., Ramsey, B., & Gibson, R. L. (2002). *Pediatr. Pulmonol.* **34**, 91–100.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.

Engebrecht, J., Nealson, K., & Silverman, M. (1983). *Cell*. **32**, 773–781.

Evans, P. (2006). *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82.

Evans, P. R. (2011). *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292.

Fahey, R. C. (2013). *Biochim. Biophys. Acta*. **1830**, 3182–3198.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). *Nucleic Acids Res.* **42**, D222–D230.

Fox, N. K., Brenner, S. E., & Chandonia, J.-M. (2014). *Nucleic Acids Res.* **42**, D304–D309.

Franke, D. & Svergun, D. I. (2009). *J. Appl. Crystallogr.* **42**, 342–346.

Franzoni, F., Colognato, R., Galetta, F., Laurenza, I., Barsotti, M., Di Stefano, R., Bocchetti, R., Regoli, F., Carpi, A., Balbarini, A., et al. (2006). *Biomed. Pharmacother. Bioméd. Pharmacothérapie*. **60**, 453–457.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). *Nucleic Acids Res.* **31**, 3784–3788.

Genghof, D. S. (1970). *J. Bacteriol.* **103**, 475–478.

Genghof, D. S. & Van Damme, O. (1964). *J. Bacteriol.* **87**, 852–862.

Genghof, D. S. & Van Damme, O. (1968). *J. Bacteriol.* **95**, 340–344.

Genghof, D. S., Inamine, E., Kovalenko, V., & Melville, D. B. (1956). *J. Biol. Chem.* **223**, 9–17.

Giddens, S. R., Feng, Y., & Mahanty, H. K. (2002). *Mol. Microbiol.* **45**, 769–783.

Goncharenko, K. V., Vit, A., Blankenfeldt, W., & Seebeck, F. P. (2015). *Angew. Chem. Int. Ed.* 2821–2824.

Griffin, J. E., Gawronski, J. D., DeJesus, M. A., Ioerger, T. R., Akerley, B. J., & Sassetti, C. M. (2011). *PLoS Pathog*. **7**, e1002251.

Gründemann, D., Harlfinger, S., Golz, S., Geerts, A., Lazar, A., Berkels, R., Jung, N., Rubbert, A., & Schomig, E. (2005). *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5256–5261.

Hand, C. E. & Honek, J. F. (2005). *J. Nat. Prod.* **68**, 293–308.

Hanlon, D. P. (1971). *J. Med. Chem.* **14**, 1084–1087.

Han, S., Eltis, L. D., Timmis, K. N., Muchmore, S. W., & Bolin, J. T. (1995). *Science*. **270**, 976–980.

Hao, Y., Kuang, Z., Walling, B. E., Bhatia, S., Sivaguru, M., Chen, Y., Gaskins, H. R., & Lau, G. W. (2012). *Cell. Microbiol.* **14**, 401–415.

Hartman, P. E. (1990). *Methods in Enzymology*, A.N.G. Lester Packer, edited by, pp. 310–318. Academic Press.

Hassan, H. M. & Fridovich, I. (1980). *J. Bacteriol.* **141**, 156–163.

Hassett, D. J., Charniga, L., Bean, K., Ohman, D. E., & Cohen, M. S. (1992). *Infect. Immun.* **60**, 328–336.

Hassett, D. J., Ma, J. F., Elkins, J. G., McDermott, T. R., Ochsner, U. A., West, S. E., Huang, C. T., Fredericks, J., Burnett, S., Stewart, P. S., et al. (1999). *Mol. Microbiol.* **34**, 1082–1093.

Hauser, A. R., Jain, M., Bar-Meir, M., & McColley, S. A. (2011). *Clin. Microbiol. Rev.* **24**, 29–70.

Heath, H. & Wildy, J. (1958). *Biochem. J.* **68**, 407–410.

Hecht, S. M. (1999). *J. Nat. Prod.* **63**, 158–168.

He, H., Ding, Y., Bartlam, M., Sun, F., Le, Y., Qin, X., Tang, H., Zhang, R., Joachimiak, A., Liu, J., et al. (2003). *J. Mol. Biol.* **325**, 1019–1030.

Held, K., Ramage, E., Jacobs, M., Gallagher, L., & Manoil, C. (2012). *J. Bacteriol.* **194**, 6387–6389.

Hendrickson, W. A. (1985). *Trans. Am. Cryst. Assoc.* **21**, 11–21.

Hickman, A. B., Namboodiri, M. A., Klein, D. C., & Dyda, F. (1999). *Cell*. **97**, 361–369.

Houtman, J. C. D., Brown, P. H., Bowden, B., Yamaguchi, H., Appella, E., Samelson, L. E., & Schuck, P. (2007). *Protein Sci.* **16**, 30–42.

Hunter, G. (1928). *Biochem. J.* **22**, 4–10.

Hu, W., Song, H., Sae Her, A., Bak, D. W., Naowarojna, N., Elliott, S. J., Qin, L., Chen, X., & Liu, P. (2014). *Org. Lett.*

Ishikawa, Y., Israel, S. E., & Melville, D. B. (1974). *J. Biol. Chem.* **249**, 4420–4427.

Ishikawa, Y. & Melville, D. B. (1970). *J. Biol. Chem.* **245**, 5967–5973.

Isupov, M. N., Obmolova, G., Butterworth, S., Badet-Denisot, M.-A., Badet, B., Polikarpov, I., Littlechild, J. A., & Teplyakov, A. (1996). *Structure*. **4**, 801–810.

Jacob, C., Jamier, V., & Ba, L. A. (2011). *Curr. Opin. Chem. Biol.* **15**, 149–155.

Jeong, J.-H., Cha, H. J., Ha, S.-C., Rojviriya, C., & Kim, Y.-G. (2014). *Biochem. Biophys. Res. Commun.* **452**, 1098–1103.

Jocelyn, P. C. (1958). *Biochem. J.* **70**, 656–660.

Jones, G. W., Doyle, S., & Fitzpatrick, D. A. (2014). *Gene*. **549**, 161–170.

Kabsch, W. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.

Kalawy-Fansa, E. (2010). M. Sc. Thesis: Untersuchungen an Glyoxalase I / Bleomycin-Resistenz-Proteinen aus *Pseudomonas aeruginosa* PAO1. Master Thesis. MPI Dortmund.

Kanehisa, M. (2002). *Novartis Found. Symp.* **247**, 91–101; discussion 101–103, 119–128, 244–252.

Keegan, R. M. & Winn, M. D. (2008). *Acta Crystallogr. D Biol. Crystallogr.* **64**, 119–124.

Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A., & Schuck, P. (2012). *Anal. Chem.* **84**, 5066–5073.

Kelley, L. A. & Sternberg, M. J. E. (2009). *Nat. Protoc.* **4**, 363–371.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C. (1958). *Nature*. **181**, 662–666.

Kibbe, W. A. (2007). *Nucleic Acids Res.* **35**, W43–W46.

Kogan, A., Gdalevsky, G. Y., Cohen-Luria, R., Goldgur, Y., Phillips, R. S., Parola, A. H., & Almog, O. (2009). *BMC Struct. Biol.* **9**, 65.

Koley, D., Ramsey, M. M., Bard, A. J., & Whiteley, M. (2011). *Proc. Natl. Acad. Sci.* **108**, 19996–20001.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., & Svergun, D. I. (2003). *J. Appl. Crystallogr.* **36**, 1277–1282.

Kozin, M. B. & Svergun, D. I. (2001). *J. Appl. Crystallogr.* **34**, 33–41.

Krissinel, E. & Henrick, K. (2004). *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Kuspa, A., Kroos, L., & Kaiser, D. (1986). *Dev. Biol.* **117**, 267–276.

Laemmli, U. K. (1970). *Nature*. **227**, 680–685.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., et al. (2007). *Bioinformatics*. **23**, 2947–2948.

Lau, G. W., Ran, H., Kong, F., Hassett, D. J., & Mavrodi, D. (2004). *Infect. Immun.* **72**, 4275–4278.

Laurenza, I., Colognato, R., Migliore, L., Del Prato, S., & Benzi, L. (2008). *BioFactors Oxf. Engl.* **33**, 237–247.

Lee, J., Wu, J., Deng, Y., Wang, J., Wang, C., Wang, J., Chang, C., Dong, Y., Williams, P., & Zhang, L.-H. (2013). *Nat. Chem. Biol.* **9**, 339–343.

Lee, J. & Zhang, L. (2014). *Protein Cell*.

Lee, S. G., Kim, Y., Alpert, T. D., Nagata, A., & Jez, J. M. (2012). *J. Biol. Chem.* **287**, 1426–1434.

Leone, E. & Mann, T. (1951). *Nature*. **168**, 205–206.

Levitt, M. & Chothia, C. (1976). *Nature*. **261**, 552–558.

Liberati, N. T., Urbach, J. M., Miyata, S., Lee, D. G., Drenkard, E., Wu, G., Villanueva, J., Wei, T., & Ausubel, F. M. (2006). *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2833–2838.

Long, F., Vagin, A. A., Young, P., & Murshudov, G. N. (2007). *Acta Crystallogr. D Biol. Crystallogr.* **64**, 125–132.

Look, D. C., Stoll, L. L., Romig, S. A., Humlicek, A., Britigan, B. E., & Denning, G. M. (2005). *J. Immunol. Baltim. Md 1950*. **175**, 4017–4023.

Lorenz, M. C. & Fink, G. R. (2002). *Eukaryot. Cell*. **1**, 657–662.

Lowther, W. T. & Matthews, B. W. (2000). *Biochim. Biophys. Acta*. **1477**, 157–167.

Lyczak, J. B., Cannon, C. L., & Pier, G. B. (2002). *Clin. Microbiol. Rev.* **15**, 194–222.

Malki, L., Yanku, M., Borovok, I., Cohen, G., Mevarech, M., & Aharonowitz, Y. (2009). *J. Bacteriol.* **191**, 5196–5204.

Mao, F., Dam, P., Chou, J., Olman, V., & Xu, Y. (2009). *Nucleic Acids Res.* **37**, D459–D463.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., et al. (2015). *Nucleic Acids Res.* **43**, D222–D226.

Markova, N. G., Karaman-Jurukovska, N., Dong, K. K., Damaghi, N., Smiles, K. A., & Yarosh, D. B. (2009). *Free Radic. Biol. Med.* **46**, 1168–1176.

Martin, T. W., Dauter, Z., Devedjiev, Y., Sheffield, P., Jelen, F., He, M., Sherman, D. H., Otlewski, J., Derewenda, Z. S., & Derewenda, U. (2002). *Structure*. **10**, 933–942.

Maruyama, M., Kumagai, T., Matoba, Y., Hayashida, M., Fujii, T., Hata, Y., & Sugiyama, M. (2001). *J. Biol. Chem.* **276**, 9992–9999.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Mavrodi, D. V., Blankenfeldt, W., & Thomashow, L. S. (2006). *Annu. Rev. Phytopathol.* **44**, 417–445.

Mavrodi, D. V., Peever, T. L., Mavrodi, O. V., Parejko, J. A., Raaijmakers, J. M., Lemanceau, P., Mazurier, S., Heide, L., Blankenfeldt, W., Weller, D. M., et al. (2010). *Appl. Environ. Microbiol.* **76**, 866–879.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). *J. Appl. Crystallogr.* **40**, 658–674.

McFadden, B. A. & Purohit, S. (1977). *J. Bacteriol.* **131**, 136–144.

McManus, I. R. (1962). *J. Biol. Chem.* **237**, 1207–1211.

Melville, D. B. & Eich, S. (1956). *J. Biol. Chem.* **218**, 647–651.

Melville, D. B., Eich, S., & Ludwig, M. L. (1957). *J. Biol. Chem.* **224**, 871–877.

Melville, D. B., Horner, W. H., & Lubschez, R. (1954). *J. Biol. Chem.* **206**, 221–228.

Mentel, M., Ahuja, E. G., Mavrodi, D. V., Breinbauer, R., Thomashow, L. S., & Blankenfeldt, W. (2009). *ChemBioChem*. **10**, 2295–2304.

Mittal, R., Aggarwal, S., Sharma, S., Chhibber, S., & Harjai, K. (2009). *FEMS Immunol. Med. Microbiol.* **57**, 156–164.

Mollica, A., Locatelli, M., Stefanucci, A., & Pinnen, F. (2012). *Molecules*. **17**, 6083–6099.

Moriarty, N. W., Grosse-Kunstleve, R. W., & Adams, P. D. (2009). *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080.

Motohashi, N. & Mori, I. (1986). *J. Inorg. Biochem.* **26**, 205–212.

Motohashi, N., Mori, I., & Sugiura, Y. (1976). *Chem. Pharm. Bull. (Tokyo)*. **24**, 2364–2368.

Motohashi, N., Mori, I., Sugiura, Y., & Tanaka, H. (1974). *Chem. Pharm. Bull. (Tokyo)*. **22**, 654–657.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367.

Nadzirin, N., Gardiner, E. J., Willett, P., Artymiuk, P. J., & Firdaus-Raih, M. (2012). *Nucleic Acids Res.* **40**, W380–W386.

Nakamura, T., Sugiura, S., Kobayashi, D., Yoshida, K., Yabuuchi, H., Aizawa, S., Maeda, T., & Tamai, I. (2007). *Pharm. Res.* **24**, 1628–1635.

Nixon, G. M., Armstrong, D. S., Carzino, R., Carlin, J. B., Olinsky, A., Robertson, C. F., & Grimwood, K. (2001). *J. Pediatr.* **138**, 699–704.

Nocek, B., Cuff, M., Evdokimova, E., Edwards, A., Joachimiak, A., & Savchenko, A. (2006). *Proteins Struct. Funct. Bioinforma.* **63**, 1102–1105.

Oda, K., Matoba, Y., Noda, M., Kumagai, T., & Sugiyama, M. (2010). *J. Biol. Chem.* **285**, 1446–1456.

Oinonen, C. & Rouvinen, J. (2000). *Protein Sci. Publ. Protein Soc.* **9**, 2329–2337.

O'Malley, Y. Q., Reszka, K. J., Spitz, D. R., Denning, G. M., & Britigan, B. E. (2004). *Am. J. Physiol. Lung Cell. Mol. Physiol.* **287**, L94–L103.

Papaioannou, E., Utari, P. D., & Quax, W. J. (2013). *Int. J. Mol. Sci.* **14**, 19309–19340.

Park, C.-H., Park, C.-H., Lee, Y.-J., Lee, S.-Y., Oh, H.-B., & Lee, J.-W. (2011). *Bull. Korean Chem. Soc.* **32**, 524–530.

Parsons, J. F., Calabrese, K., Eisenstein, E., & Ladner, J. E. (2003). *Biochemistry*. **42**, 5684–5693.

Parsons, J. F., Calabrese, K., Eisenstein, E., & Ladner, J. E. (2004a). *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2110–2113.

Parsons, J. F., Song, F., Parsons, L., Calabrese, K., Eisenstein, E., & Ladner, J. E. (2004b). *Biochemistry*. **43**, 12427–12435.

Paul, B. & Snyder, S. (2010). *Cell Death Differ.* **17**, 1134–1140.

Pei, J., Kim, B.-H., & Grishin, N. V. (2008). *Nucleic Acids Res.* **36**, 2295–2300.

Peltekova, V. D., Wintle, R. F., Rubin, L. A., Amos, C. I., Huang, Q., Gu, X., Newman, B., Oene, M. V., Cescon, D., Greenberg, G., et al. (2004). *Nat. Genet.* **36**, 471–475.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., & Svergun, D. I. (2012). *J. Appl. Crystallogr.* **45**, 342–350.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Pfeiffer, C., Bauer, T., Surek, B., Schömig, E., & Gründemann, D. (2011). *Food Chem.* **129**, 1766–1769.

Pierson, L. S., Gaffney, T., Lam, S., & Gong, F. (1995). *FEMS Microbiol. Lett.* **134**, 299–307.

Pierson, L. S. & Thomashow, L. S. (1992). *Mol. Plant-Microbe Interact. MPMI*. **5**, 330–339.

Popp, M. (2013). M. Sc. Thesis: Charakterisierung von PA3127 und PA3128 aus *Pseudomonas aeruginosa* PAO1. Universität Bayreuth.

Putnam, D. K., Lowe, E. W., & Meiler, J. (2013). *Comput. Struct. Biotechnol. J.* **8**,.

Rada, B. & Leto, T. L. (2013). *Trends Microbiol.* **21**, 73–81.

Rahman, I., Gilmour, P. S., Jimenez, L. A., Biswas, S. K., Antonicelli, F., & Aruoma, O. I. (2003). *Biochem. Biophys. Res. Commun.* **302**, 860–864.

Recinos, D. A., Sekedat, M. D., Hernandez, A., Cohen, T. S., Sakhtah, H., Prince, A. S., Price-Whelan, A., & Dietrich, L. E. P. (2012). *Proc. Natl. Acad. Sci.* **109**, 19420–19425.

Reszka, K. J., O'Malley, Y., McCormick, M. L., Denning, G. M., & Britigan, B. E. (2004). *Free Radic. Biol. Med.* **36**, 1448–1459.

Rife, C. L., Pharris, R. E., Newcomer, M. E., & Armstrong, R. N. (2002). *J. Am. Chem. Soc.* **124**, 11001–11003.

Rigsby, R. E., Rife, C. L., Fillgrove, K. L., Newcomer, M. E., & Armstrong, R. N. (2004). *Biochemistry*. **43**, 13666–13673.

Rupp, B. (2009). Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology New York: Garland Science.

Salt, H. B. (1931). *Biochem. J.* **25**, 1712–1719.

Sao Emani, C., Williams, M. J., Wiid, I. J., Hiten, N. F., Viljoen, A. J., Pietersen, R.-D. D., van Helden, P. D., & Baker, B. (2013). *Antimicrob. Agents Chemother.* **57**, 3202–3207.

Sasikaran, J., Ziemski, M., Zadora, P. K., Fleig, A., & Berg, I. A. (2014). *Nat. Chem. Biol.* **10**, 371–377.

Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363.

Seebeck, F. P. (2010). *J. Am. Chem. Soc.* **132**, 6632–6633.

Sillitoe, I., Cuff, A. L., Dessailly, B. H., Dawson, N. L., Furnham, N., Lee, D., Lees, J. G., Lewis, T. E., Studer, R. A., Rentzsch, R., et al. (2013). *Nucleic Acids Res.* **41**, D490–D498.

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S. L., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., et al. (2000). *Nature*. **406**, 959–964.

Strelko, C. L., Lu, W., Dufort, F. J., Seyfried, T. N., Chiles, T. C., Rabinowitz, J. D., & Roberts, M. F. (2011). *J. Am. Chem. Soc.* **133**, 16386–16389.

Sugiyama, M., Kumagai, T., Hayashida, M., Maruyama, M., & Matoba, Y. (2002). *J. Biol. Chem.* **277**, 2311–2320.

Sukdeo, N., Clugston, S. L., Daub, E., & Honek, J. F. (2004). *Biochem. J.* **384**, 111.

Sukdeo, N. & Honek, J. F. (2007). *Biochim. Biophys. Acta BBA - Proteins Proteomics*. **1774**, 756–763.

Svergun, D., Barberato, C., & Koch, M. H. J. (1995). *J. Appl. Crystallogr.* **28**, 768–773.

Svergun, D. I. (1992). *J. Appl. Crystallogr.* **25**, 495–503.

Tanner, K. G., Langer, M. R., & Denu, J. M. (2000a). *Biochemistry*. **39**, 11961–11969.

Tanner, K. G., Langer, M. R., Kim, Y., & Denu, J. M. (2000b). *J. Biol. Chem.* **275**, 22048–22055.

Tanret, C. (1909). *Compt. Rend. Acad. Sci.* **49**, 222–224.

Ta, P., Buchmeier, N., Newton, G. L., Rawat, M., & Fahey, R. C. (2011). *J. Bacteriol.* **193**, 1981–1990.

Taubert, D., Jung, N., Goeser, T., & Schömig, E. (2009). *Gut*. **58**, 312–314.

Trievel, R. C., Rojas, J. R., Sterner, D. E., Venkataramani, R. N., Wang, L., Zhou, J., Allis, C. D., Berger, S. L., & Marmorstein, R. (1999). *Proc. Natl. Acad. Sci.* **96**, 8931–8936.

Usher, L. R., Lawson, R. A., Geary, I., Taylor, C. J., Bingle, C. D., Taylor, G. W., & Whyte, M. K. B. (2002). *J. Immunol. Baltim. Md 1950*. **168**, 1861–1868.

Vetting, M. W., Magnet, S., Nieves, E., Roderick, S. L., & Blanchard, J. S. (2004). *Chem. Biol.* **11**, 565–573.

Vit, A. (2010). M. Sc. Thesis: Charakterisierung von Enzymen aus der Ergothioneinbiosynthese. TU Dortmund.

Vit, A., Misson, L., Blankenfeldt, W., & Seebeck, F. P. (2014). *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **70**, 676–680.

Vit, A., Misson, L., Blankenfeldt, W., & Seebeck, F. P. (2015). *ChemBioChem*. **16**, 119–125.

Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Crystallogr.* **36**, 860–864.

Wang, Y., Kern, S. E., & Newman, D. K. (2010). *J. Bacteriol.* **192**, 365–369.

Wang, Y. & Newman, D. K. (2008). *Environ. Sci. Technol.* **42**, 2380–2386.

Wargo, M. J. (2013). *Appl. Environ. Microbiol.* **79**, 2112–2120.

Weber, I. (2011). B. Sc. Thesis: Biochemische und strukturelle Untersuchungen zu Glyoxalase I/Bleomycinresistenzproteinen aus *Pseudomonas aeruginosa* PAO1. Universität Bayreuth.

Weinert, T., Olieric, V., Waltersperger, S., Panepucci, E., Chen, L., Zhang, H., Zhou, D., Rose, J., Ebihara, A., Kuramitsu, S., et al. (2015). *Nat. Methods*. **12**, 131–133.

Weiss, M. S. (2001). *J. Appl. Crystallogr.* **34**, 130–135.

Wilson, R., Sykes, D. A., Watson, D., Rutman, A., Taylor, G. W., & Cole, P. J. (1988). *Infect. Immun.* **56**, 2515–2517.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., et al. (2011). *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242.

Winson, M. K., Camara, M., Latifi, A., Foglino, M., Chhabra, S. R., Daykin, M., Bally, M., Chapon, V., Salmond, G. P., & Bycroft, B. W. (1995). *Proc. Natl. Acad. Sci.* **92**, 9427–9431.

Winsor, G. L., Lam, D. K. W., Fleming, L., Lo, R., Whiteside, M. D., Yu, N. Y., Hancock, R. E. W., & Brinkman, F. S. L. (2011). *Nucleic Acids Res.* **39**, D596–D600.

Wriggers, W. (2012). *Acta Crystallogr. D Biol. Crystallogr.* **68**, 344–351.

Yu, S. (2009). PhD Thesis: Assessment of *Pseudomonas* Quinolone Signal Response Protein PqsE and Preliminary Functional Annotation of Hypothetical Protein PA0803 from *Pseudomonas aeruginosa*. MPI Dortmund.

Yu, S., Vit, A., Devenish, S., Mahanty, H. K., Itzen, A., Goody, R. S., & Blankenfeldt, W. (2011). *BMC Struct. Biol.* **11**, 33.

Zhao, Q., Wang, M., Xu, D., Zhang, Q., & Liu, W. (2015). *Nature*. **518**, 115–119.

Zwart, P. H., Grosse-Kunstleve, R. W., & Adams, P. D. (2005). *CCP4 Newsl.* **Winter**, Contribution 7.

16 Appendices

16.1 Additional Plasmids Generated in This Study

Additional plasmids generated during this study are listed in Table 45.

Table 45: Additional plasmids generated during this study or related studies under the supervision of the author.

Plasmid	restrictions sites	Host	Source
pET19m_PA0817	NdeI / XhoI	*E. coli*	Weber, 2011
pET19m_PA0865	NdeI / BamHI	*E. coli*	this study
pET19m_PA1359	NdeI / XhoI	*E. coli*	this study
pET19m_PA2024	NdeI / XhoI	*E. coli*	this study
pET19m_PA2546	NdeI / BamHI	*E. coli*	this study
pET19m_PA2721	NdeI / XhoI	*E. coli*	this study
pET19m_PA3127_1-257	NdeI / BamHI	*E. coli*	this study
pET19m_PA3128	NdeI / XhoI	*E. coli*	this study
pET19m_PA3389	NdeI / XhoI	*E. coli*	this study
pET19m_PA3505	NdeI / XhoI	*E. coli*	this study
pET19m_PA3956	NdeI / XhoI	*E. coli*	this study
pET19m_PA4184	NdeI / XhoI	*E. coli*	this study

16.2 Additional Datasets Processed and Structures Refined in This Study

The crystal structures of PA0803, PA1353, PA3127-NTD and PA4641 have been determined in studies by Yu, 2009, Kalawy-Fansa, 2010, and Popp, 2013. The crystallographic data obtained at that time was reprocessed with current software versions of *XDS* and *Aimless* and the crystal structures were rebuild and re-refined in the course of this study. This procedure resulted in optimized structural models as well as improved data collection, processing and refinement statistics and. These statistics are given in Table 46, Table 47, Table 48 and Table 49.

Table 46: Data collection statistics for *P. aeruginosa* βαβββ module containing proteins PA0803 and PA1353. Data were collected in studies by Yu, 2009, and Kalawy-Fansa, 2010. Values in parentheses are for the highest resolution shell. PYO = pyocyanin

Dataset	PA0803 Se SAD[+]	PA0803 apo	PA0803 PYO	PA1353 apo
Wavelength (Å) / beamline[≠]	0.9792 / SLS, X10SA (PXII)	1.5418/ MPI Dortmund	0.9184 / BESSY II, BL14.1	0.9787 / SLS, X10SA (PXII)
Resolution range (Å)	19.84-1.95 (2.0-1.95)	46.55-1.62 (1.65-1.62)	46.39-1.71 (1.74-1.71)	49.11-1.75 (1.78-1.75)
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12_1$
Unit cell parameters (Å)	85.35 101.6 92.57	84.91 102.1 93.11	84.26 101.8 92.77	64.31 66.54 76.07
(°)	90 90 90	90 90 90	90 90 90	90 90 90
Mosaicity (°)[†]	0.234	0.211	0.109	0.286
Total No. of measured reflections	622550 (41972)	1258371 (12836)	357368 (19052)	387835 (20979)
Unique reflections	58285 (4029)	102661 (4811)	86649 (4588)	33644 (1820)
Multiplicity	5.6 (5.4)	12.3 (2.7)	4.1 (4.2)	11.5 (11.5)
Mean I/σ(I)	31.0 (11.8)	14.9 (3.5)	11.6 (2.1)	17.1 (2.1)
Completeness (%)	98. 3 (98)	99.4 (95.4)	99.9 (99.9)	100 (100)
R_{meas} (%)[‡]	5.2 (17.7)	15.7 (37.6)	9.1 (81.2)	7.9 (122.4)
R_{pim} (%)[$]	2.4 (8.4)	3.5 (21.9)	4.5 (39.3)	2.3 (35.8)

[+] Data collection statistics for unmerged Friedel pairs

[≠] SLS: Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland), BESSY: Berlin Electron Storage Ring Society for Synchrotron Radiation (Helmholtz Zentrum Berlin, Berlin, Germany)

[†] Mosaicity values reported by *XDS* (Kabsch, 2010)

[‡] $R_{meas} = \Sigma_{hkl} (N/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - < I(hkl) > \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where N is the number of observations of the reflection with index hkl and I_i is the intensity of its i^{th} observation.

[$] $R_{pim} = \Sigma_{hkl} (1/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - < I(hkl) > \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$ (Weiss, 2001)

Table 47: Refinement statistics for *P. aeruginosa* βαβββ module containing proteins PA0803 and PA1353. Data were collected in studies by Yu, 2009, and Kalawy-Fansa, 2010. Values in parentheses are for the highest resolution shell. PYO = pyocyanin

Dataset	PA0803 apo	PA0803 PYO	PA1353 apo
Resolution range (Å)	46.55-1.62 (1.65-1.62)	46.39-1.71 (1.74-1.71)	49.11-1.75 (1.78-1.75)
R_{work} (%)	17.75 (35.51)	15.37 (26.28)	17.43 (23.84)
R_{free} (%)	20.20 (38.47)	18.45 (31.45)	21.58 (28.48)
No. of non-H atoms			
Protein	4430	5282	2119
Ion	-	10	-
Ligand	52	80	-
Water	705	799	249
R.m.s. deviations			
Bonds (Å)	0.005	0.008	0.020
Angles (°)	0.999	1.166	1.818
Average B factors ($Å^2$)			
Protein	27	23	33
Ion	-	50	-
Ligand	53	38	-
Water	40	38	42
Ramachandran plot			
Favored regions (%)	98.1	97.8	98.2
Outliers (%)	0	0.2	0.4
MolProbity score[#]	1.19	1.17	1.09
PDB entry code	-	-	-

[#] As reported by *MolProbity* (Chen *et al.*, 2010)

Table 48: Data collection statistics for *P. aeruginosa* βαβββ module containing proteins PA1353, PA3127-NTD and PA4641. Data were collected in studies by Kalawy-Fansa, 2010 and Popp, 2013. Values in parentheses are for the highest resolution shell. PYO = pyocyanin

Dataset	PA1353 PYO	PA3127-NTD apo	PA4641 apo	PA4641 PYO
Wavelength (Å) / beamline≠	0.9787 / SLS, PXII	0.9184 / BESSY II, BL14.1	0.9787 / SLS, PXII	0.9787 / SLS, PXII
Resolution range (Å)	46.57-1.45 (1.47-1.45)	38.12-1.55 (1.58-1.55)	44.79-1.95 (2.0-1.95)	48.90-1.85 (1.89-1.85)
Space group	$P2_12_12_1$	P1	P1	P1
Unit cell parameters (Å)	65.58 66.13 76.05	43.38 43.6 66.2	30.96 49.32 86.69	30.95 49.24 86.97
(°)	90 90 90	89.92 78.41 61.6	96.39 91.25 91.81	96.46 91.28 91.99
Mosaicity (°)†	0.133	0.142	0.242	0.095
Total No. of measured reflections	488313 (23926)	215582 (10142)	97274 (6923)	88049 (5619)
Unique reflections	59343 (2908)	58424 (2829)	35220 (2465)	40844 (2526)
Multiplicity	8.2 (8.2)	3.7 (3.6)	2.8 (2.8)	2.2 (2.2)
Mean I/σ(I)	20.6 (2.4)	16.9 (2.2)	6.9 (2.2)	9.6 (2.3)
Completeness (%)	100 (100)	96.9 (95.3)	94.8 (95.6)	93.9 (94.7)
R_{meas} (%)‡	5.8 (102.1)	4.7 (66)	10.4 (49.2)	6.9 (54.1)
R_{pim} (%)$	2.0 (35.4)	2.4 (34.6)	6.0 (28.0)	4.3 (34.8)

≠ SLS: Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland), BESSY: Berlin Electron Storage Ring Society for Synchrotron Radiation (Helmholtz Zentrum Berlin, Berlin, Germany)

† Mosaicity values reported by *XDS* (Kabsch, 2010)

‡ $R_{meas} = \Sigma_{hkl} (N/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where N is the number of observations of the reflection with index hkl and I_i is the intensity of its i^{th} observation.

\$ $R_{pim} = \Sigma_{hkl} (1/(N-1))^{1/2} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl)$ (Weiss, 2001).

Table 49: Refinement statistics for *P. aeruginosa* βαβββ module containing proteins PA1353, PA3127-NTD and PA4641. Data were collected in studies by Kalawy-Fansa, 2010 and Popp, 2013. Values in parentheses are for the highest resolution shell. PYO = pyocyanin

Dataset	PA1353 PYO	PA3127-NTD apo	PA4641 apo	PA4641 PYO
Resolution range (Å)	46.57-1.45	38.12-1.55	44.79-1.95	48.90-1.85
	(1.47-1.45)	(1.58-1.55)	(2.0-1.95)	(1.89-1.85)
R_{work} (%)	16.24 (24.68)	17.05 (24.2)	18.6 (23.38)	16.48 (24.14)
R_{free} (%)	18.93 (23.9)	19.41 (28.9)	22.47 (27.98)	20.28 (32.33)
No. of non-H atoms				
Protein	2184	3506	4245	4252
Ion	-	-	-	-
Ligand	32	14	-	64
Water	295	330	238	274
R.m.s. deviations				
Bonds (Å)	0.016	0.008	0.0056	0.006
Angles (°)	1.534	1.248	0.950	0.957
Average B factors (Å^2)				
Protein	26	25	33	30
Ion	-	-	-	-
Ligand	26	32	-	27
Water	38	35	42	39
Ramachandran plot				
Favored regions (%)	98.6	97.3	96.5	97.4
Outliers (%)	0	0	0	0.2
MolProbity score[#]	1.14	0.95	1.32	0.93
PDB entry code	-	-	-	-

[#] As reported by *MolProbity* (Chen *et al.*, 2010)

Danksagung

Diese Arbeit wurde unter der Betreuung durch Prof. Dr. Wulf Blankenfeldt am Lehrstuhl für Biochemie der Universität Bayreuth, Deutschland, sowie in der Abteilung *Struktur und Funktion der Proteine* am Helmholtz-Zentrum für Infektionsforschung in Braunschweig, Deutschland, im Zeitraum von Januar 2011 bis März 2015 angefertigt. Das Ergothioneinbiosynthese-Projekt ist eine Zusammenarbeit mit der Arbeitsgruppe von Prof. Dr. Florian P. Seebeck von der Universität Basel in Basel, Schweiz.

Hiermit möchte ich den Menschen danken, die mich in der Zeit der Doktorarbeit fachlich und persönlich unterstützt haben und somit diese Doktorarbeit ermöglicht haben:

Prof. Dr. Wulf Blankenfeldt, für das interessante und vielschichtige Projekt, für das Vertrauen in mich und meine Arbeit, den wissenschaftlichen Freiraum, den er mir gegeben hat, sowie die Unterstützung bei allen wissenschaftlichen Vorhaben,
Prof. Dr. Florian P. Seebeck, als meinen wissenschaftsbegeisterten Kooperationspartner im Ergothioneinbiosynthese-Projekt , der mir das Gefühl gibt, entscheidend zu seinen Projekten beigetragen zu haben,
Prof. Dr. Michael Steinert, für die Übernahme der Zweitkorrektur,
PD Dr. Michael Hust, für die Übernahme des Vorsitzes meiner Promotionskommission,
den Mitgliedern meines *thesis committee* **Prof. Dr. Susanne Häußler**, unserer Kooperationspartnerin, für die *P. aeruginosa* Stämme, und **Dr. Andrea Scrima** für hilfreiche Diskussionen und wirklich guten Humor,
Dr. Michael Weyand und **Dr. Sebastién Moniot,** sowie allen **Mitgliedern der Strukturbiologie** der Universität Bayreuth und des Helmholtz-Zentrums für Infektionsforschung für die Bereitschaft, all meine kristallographischen Fragen zu beantworten und Hilferufe zu erhören, sowie für Datensammlungen der Kristalle,
meiner Master-Studentin und Kollegin im βαβββ-Modulproteine-Projekt **Monika Popp** für ihr Interesse und Ihre tatkräftige Unterstützung, sowie meinen Bachelor-Studentinnen **Irina Weber** und **Julia Bittner** für ihre helfenden Hände, und meinen Vorgängern auf diesem Projekt, **Dr. Shen Yu** und **Eyad Kalawy-Fansa**,

all meinen Kollegen und ehemaligen Kollegen, vor allem **Dr. Emerich-Mihai Gazdag**, **Christina Diederich**, **Christian Feiler**, **Tobias Bock**, **Nina Schwemmlein** und **Florian Witzgall** für gute wissenschaftliche Diskussionen und eine tolle Atmosphäre,

allen weiteren Mitgliedern und Ehemaligen der AG Steegborn, vor allem **Silke Kleinbölting**, als meine kompetente Bürokollegin, **Prof. Dr. Clemens Steegborn**, **Dr. Melanie Fischer**, **Dr. Frank Fischer**, **Norbert Grillenbeck**, **Edith Guthmann, Lisa Meisel**, **Susanne Schäfer**, **Renate Crowe, Martin Pannek**, **Benjamin Sünkel** und **Gabriele Kassler** für wissenschaftliche, technische und persönliche Unterstützung, sowie die tolle Atmosphäre, die Maiwanderungen und alle spontanen Grill-Aktionen,

den Doktoranden von Prof. Dr. Florian Seebeck, **Laëtitia Misson**, **Kristina V. Goncharenko, Matthias Knop** und **Gabriel T. Mashabela** für die erfolgreiche Zusammenarbeit und ein Bad im Rhein in Basel,

allen **Mitgliedern der Abteilung SFPR**, im Besonderen den Gruppen **SBAU** und **RPEX**, vor allem aber **Dr. Caroline Behrens** für hilfreiche Kommentare und Korrekturen zu meiner Doktorarbeit, und **Milica Bajagić**, **Archna Gautam** und **Pooja Sadana** für das gemeinsame Testen vieler Restaurants in Braunschweig,

Ulrike Beutling und **Prof. Dr. Mark Brönstrup** für die MS-Experimente zu PA3127, **Dr. Joop van den Heuvel** für Zugang zur ITC und **Prof. Dr. Susanne Engelmann** für Zutritt zum S2 Labor, sowie **Dr. Jörn Krauße** für Unterstützung bei der Analyse der SAXS Daten, **Prof. Dr. Lars Dietrich** für *P. aeruginosa* PA14 und **Prof. Dr. Rolf Breinbauer** für Pyocyanin,

der **Helmholtz Graduate School for Infection Research** für finanzielle Unterstützung und die vielen Weiterbildungsangebote,

dem **BESSY II** (HZB, Berlin), **SLS** (PSI, Villigen, Schweiz), **PETRA III** (DESY, EMBL, Hamburg) und **ESRF** (Grenoble, Frankreich) für Zugang zu ihren Anlagen und Einrichtungen, und ihren **Wissenschaftlern** für Unterstützung bei der Datensammlung,

meiner Schwester Allessia und **meinen Eltern Gabi und Franco,** die immer für mich da waren, mich unterstützt und motiviert haben,

und **Christian** für die Unterstützung, das Verständnis und die Liebe, die er mir entgegenbringt, und all die Abenteuer des Lebens, die wir gemeinsam erleben und erleben werden.

Curriculum Vitae

Name	Katharina Allegra Vit
Geburtstag und -ort	20. September 1985 in Bochum, Deutschland
Nationalitäten	Deutsch und Italienisch

Akademischer Werdegang

Januar 2011 – März 2015

Promotion in der Arbeitsgruppe von Prof. Dr. Wulf Blankenfeldt am Lehrstuhl Biochemie der Universität Bayreuth, Bayreuth und am Helmholtz-Zentrum für Infektionsforschung, Braunschweig

Januar – Dezember 2010

Masterarbeit in den Arbeitsgruppen von Dr. Florian P. Seebeck und Dr. Wulf Blankenfeldt, Abteilung Prof. Dr. Roger S. Goody, am Max-Planck-Institut für Molekulare Physiologie, Dortmund

Oktober 2008 – Dezember 2010

Master of Science in Chemischer Biologie, TU Dortmund, Dortmund

März – September 2008

Bachelorarbeit in der Arbeitsgruppe von Dr. Susanne Eschenburg, Abteilung Prof. Dr. Alfred Wittinghofer, am Max-Planck-Institut für Molekulare Physiologie, Dortmund

Oktober 2005 – September 2008

Bachelor of Science in Chemischer Biologie, TU Dortmund, Dortmund

Schulischer Werdegang

Juni 2005

Bilinguales Abitur Deutsch/Englisch am Hellweg-Gymnasium, Bochum-Wattenscheid

Juli 2003 – Juni 2005

Bilinguale Oberstufe Deutsch/Englisch am Hellweg-Gymnasium, Bochum-Wattenscheid

August 2002 – Juni 2003

Educandato Statale Collegio Uccellis – Liceo Classico Europeo, Udine, Italien

August 1996 – Juli 2002

Bilingualer Zweig Deutsch/Englisch am Hellweg-Gymnasium, Bochum-Wattenscheid

August 1992 – Juli 1996

Dietrich-Boenhoffer-Grundschule, Bochum-Wattenscheid

Stipendien und Mitgliedschaften

Mai 2012 – heute

Mitglied der Gesellschaft für Molekularbiologie und Biochemie (GBM)

Oktober 2013 – März 2015

Mitglied der *International Graduate School for Infection Research* des Helmholtz-Zentrums für Infektionsforschung, Braunschweig

November 2013

Scholarship für *Falling Walls Lab and Conference* 2013 in Berlin

Vollständige Publikationsliste

A. Vit, G. T. Mashabela, W. Blankenfeldt, F. P. Seebeck: Structure of the Ergothioneine-Biosynthesis Amidohydrolase EgtC. *ChemBioChem*. **2015** Jul 6; 16(10): 1490-1496.

A. Vit*, K. V. Goncharenko*, W. Blankenfeldt, F. P. Seebeck: Structure of the Sulfoxide Synthase EgtB from the Ergothioneine Biosynthetic Pathway. *Angew. Chem. Int. Ed. Engl*. **2015** Feb 23; 54(9): 2821-2824.

A. Vit*, L. Misson*, W. Blankenfeldt, F. P. Seebeck: Ergothioneine Biosynthetic Methyltransferase EgtD Reveals the Structural Basis of Aromatic Amino Acid Betaine Biosynthesis. *ChemBioChem*. **2015** Jan 2; 16(1): 119-125.

A. Vit, L. Misson, W. Blankenfeldt, F. P. Seebeck: Crystallization and preliminary X-ray analysis of the ergothioneine-biosynthetic methyltransferase EgtD. *Acta Crystallogr F Struct Biol Commun*. **2014** May; 70(Pt 5): 676-680.

S. Yu, A. Vit, S. Devenish, H. K. Mahanty, A. Itzen, R. S. Goody, W. Blankenfeldt: Atomic resolution structure of EhpR: phenazine resistance in *Enterobacter agglomerans* Eh1087 follows principles of bleomycin/mitomycin C resistance in other bacteria. *BMC Struct. Biol.* **2011** Aug 17; 11: 33.

**Diese Autoren haben in gleichem Maße zu der Veröffentlichung beigetragen.*